# Étude de la variabilité de réponse immunitaire innée chez l'Homme : une approche évolutive et moléculaire

Matthieu Deschamps

# THÈSE DE DOCTORAT
# DE L'UNIVERSITÉ PIERRE ET MARIE CURIE

Spécialité
**Génétique Humaine**
École doctorale : Complexité du Vivant (ED515)

Présentée par
**Matthieu DESCHAMPS**

Pour obtenir le grade de
**DOCTEUR DE L'UNIVERSITÉ PIERRE ET MARIE CURIE**

## Étude de la variabilité de réponse immunitaire innée chez l'Homme: une approche évolutive et moléculaire

Thèse dirigée par Lluís QUINTANA-MURCI

Soutenue le 29 septembre 2015 devant le jury composé de :

| | |
|---|---|
| M. Dominique HIGUET | Président du jury |
| M. Alexandre ALCAIS | Rapporteur |
| M. David COMAS | Rapporteur |
| M. Matthew ALBERT | Examinateur |
| M. Luis BARREIRO | Examinateur |
| M. Lluís QUINTANA-MURCI | Directeur de thèse |

# THÈSE DE DOCTORAT
# DE L'UNIVERSITÉ PIERRE ET MARIE CURIE

Spécialité
**Génétique Humaine**
École doctorale : Complexité du Vivant (ED515)

Présentée par
**Matthieu DESCHAMPS**

Pour obtenir le grade de
**DOCTEUR DE L'UNIVERSITÉ PIERRE ET MARIE CURIE**

## Étude de la variabilité de réponse immunitaire innée chez l'Homme: une approche évolutive et moléculaire

Thèse dirigée par Lluís QUINTANA-MURCI

Soutenue le 29 septembre 2015 devant le jury composé de :

| | |
|---|---|
| M. Dominique HIGUET | Président du jury |
| M. Alexandre ALCAIS | Rapporteur |
| M. David COMAS | Rapporteur |
| M. Matthew ALBERT | Examinateur |
| M. Luis BARREIRO | Examinateur |
| M. Lluís QUINTANA-MURCI | Directeur de thèse |

# Remerciements

Comme cette partie est contradictoire ! Ecrite en dernière, elle ouvre pourtant le manuscrit. C'est l'une des sections les plus courtes, les plus habituelles et sans conteste la moins pertinente scientifiquement parlant. Pourtant, ces quelques lignes sont celles qui seront les plus lues de tout cet ouvrage, disséquées par de nombreux regards à la recherche d'originalité, de personnalité, du mot juste et de l'idée pertinente, raison pour laquelle il faut les mûrir longuement avant de les écrire et inlassablement les reformuler. En espérant ne décevoir personne.

Je tiens tout d'abord à remercier les personnes qui ont accepté de faire partie de mon jury : Dominique Higuet qui m'a fait l'honneur de le présider, Matthew Albert et Luis Barreiro qui l'ont examinée, ainsi qu'Alexandre Alcaïs et David Comas qui, en plus de juger ma soutenance, ont été rapporteurs de ma thèse.

Lluis, je sais que tu te jetteras sur cette partie à l'instant même où je te remettrai cette dernière version de mon manuscrit entre les mains. Je te remercie de m'avoir accueilli dans ton laboratoire, de m'avoir permis de faire deux projets très différents, de m'avoir donné l'opportunité de m'ouvrir à la bioinformatique, de m'avoir écouté, conseillé et soutenu dans mes démarches pour la suite. Merci pour la diplomatie qui a parfois été nécessaire et pour la confiance que tu m'as accordée. Je tiens aussi à te remercier pour la ferveur avec laquelle tu m'as transmis ta passion scientifique. Je me rappelle avoir été captivé lorsque j'avais suivi ton intervention en Master il y a quelques années de cela, profond intérêt qui ne s'est depuis jamais éteint et que tu as aidé à nourrir.

Je remercie aussi les personnes qui m'ont encadré au cours de mes différents projets et qui ont ainsi joué un rôle majeur dans le déroulement de ma thèse. Julien, qui a aidé à me canaliser dans mon projet « humide », Etienne, qui m'a mis le pied à l'étrier en génétique évolutive et Guillaume, qui m'a pris sous son aile et s'est montré particulièrement patient lorsque j'ai définitivement lâché les pipettes pour le clavier. Merci à l'ensemble des collaborateurs avec qui j'ai eu la chance de travailler au cours de ces dernières années : Olivier Neyrolles, Geanncarlo Lugo et Ludovic Tailleux, sans qui cette thèse n'aurait pas été possible. Merci aux membres de mon comité de thèse pour le regard critique qu'ils ont porté sur mon travail et leur souci du bon déroulement de ma thèse.

Merci à l'ensemble des membres du laboratoire pour l'esprit qui y règne.

Hélène, chère Maître Jedi à l'alcool déshydrogénase défaillante, merci pour les discussions plus ou moins scientifiques, constructives et/ou sérieuses que nous avons eues, pour tous les cookies que tu as rapportés, les mondialement connues « GEH Garden Parties », toutes les franches rigolades que nous avons partagées. J'espère que je me montrerai à la hauteur de ton organisation légendaire dans le futur (malgré mes efforts, il me reste du chemin à parcourir pour arriver à ton niveau).

Christine, discrète et pourtant si attentive au bien-être de chacun, merci pour tes attentions bienveillantes, qu'elles soient professionnelles (oui, je suis bien installé, le tabouret n'est pas trop haut et la hôte fonctionne bien) ou personnelles (je sais que j'ai fait le bon choix pour mon orientation future.

Maud, qui recrute si efficacement pour le volley que l'armée américaine compte la débaucher sous peu, merci pour ton aide précieuse qui m'a permis de ne pas fracasser mon

écran un nombre significatif de fois (mais je me suis peut être trompé de test statistique, il faudra que je voie ça avec toi) et pour les discussions politiques que nous avons eues. J'ai encore des progrès à faire pour suivre celles que tu as avec Guillaume mais qui sait, peut-être un jour y arriverai-je ?

Maxime, qui a invariablement dû m'attendre 45 minutes à tous les repas, merci de m'avoir montré que la plupart des problèmes que je rencontrais en bioinformatique résultaient d'une mauvaise formulation de la question que je posais (et de m'avoir aidé à trouver la solution le cas échéant).

Nora, qui connaît tout le monde à l'Institut Pasteur, merci (je suis bref parce que j'ai peur de dire une chose qu'il ne fallait pas et de recevoir un *high kick* en représailles). Avant que je ne parte, il faudra que nous discutions de la stratégie de *gating* pour le FACS, deux ou trois points ne sont pas encore pas tout à fait clairs je pense (c'est à ce moment là que je m'enfuie en courant).

Marie, Fée du Cluster, digne héritière capillaire de ton mentor du sixième, merci pour tes conseils concernant l'électro-jazz, ta bonne humeur permanente et les discussions scientifiques des PhD meetings pour lesquelles je remercie aussi Maud, Katie, Lucas et Eric.

Etienne le Maître des Forces Obscures aux goûts musicaux douteux et Guillaume le Grand Rugbyman des Temps Passés (pardon Mademoiselle), l'inséparable paire des labmeetings que je me dois de citer de nouveau pour l'aide qu'ils m'ont apportée, leur disponibilité, la pertinence de leurs remarques scientifique et la (non) pertinence de leurs remarques non scientifiques, merci.

Merci aussi à Barbara, indéniablement la personne la plus discrète du sixième étage (bon courage pour continuer à survivre avec ces zigotos).

Un très grand merci à Cécile qui, bien qu'elle soit un petit peu loin (30 mètres, c'est long à parcourir !), nous apporte un soutien administratif quotidien sans faille.

Je remercie aussi les personnes qui sont passées par le laboratoire et qui sont depuis parties vers d'autres destinations. Je pense tout d'abord à Katie, dont les travaux ont donné naissance à mon premier projet, qui m'a aidé tout au long de sa présence dans l'équipe, que j'ai embrigadée pour un marathon que je n'ai finalement pas couru et qu'elle a terminé, qui m'a montré que les Anglais n'étaient finalement pas si perfides. C'est un véritable honneur que d'avoir publié avec toi mes premiers résultats. Je pense aussi à Eddie qui restera à mes yeux un grand pédagogue qui s'ignore, à Choumouss, Stéphane…

Je me dois aussi de mentionner Jean-Louis Serre et Bruno Lemaitre qui m'ont montré ce qu'étaient la Génétique, la recherche et qui ont allumé la flamme qui a guidé l'ensemble de mon orientation universitaire. Sans eux je serai passé à côté d'une chose merveilleuse.

À tous mes amis, pour les déjeuners à Pasteur, les pauses café, les échanges constructifs sur nos impressions, nos joies, nos doutes, les courses à pied bizarres, les conseils musicaux, les apéritifs à la maison, l'ouverture d'esprit qu'ils m'ont apporté sans le savoir, les cours de langage des signes et plus généralement les excellents moments passés au cours de ces années qui sont nécessaires pour ne pas s'enfermer dans son travail, merci.

À ma famille, mes grands-mères, mes parents adorés, ma géniale petite sœur : votre écoute, votre soutien inconditionnel, votre intérêt pour mon travail alors qu'il peut vous paraître abstrait, vos encouragements incessants, votre accompagnement au cours de ces trois années ainsi qu'au cours des récentes décisions que j'ai prises… Je ne pourrai exprimer la chance que j'ai. Je n'en serai certainement pas là sans toutes ces attentions. Merci du fond du cœur.

Et finalement Marion, merci d'avoir partagé les bons moments de cette thèse, de m'avoir soutenu lors des instants plus difficiles, d'avoir supporté mon obsession pour le travail qui n'a fait que s'accentuer ces derniers mois. Merci pour la richesse de tes attentions quotidiennes qui sont trop nombreuses pour pouvoir être toutes mentionnées ici et qui font de ta présence un pilier sur lequel je peux m'appuyer pour évoluer chaque jour. Tu ne peux imaginer à quel point tu es nécessaire à mon équilibre. Je t'aime.

# Table des matières

# Table des figures

# Liste des abréviations

| | |
|---|---|
| ΔDAF | Différence de fréquences de l'allèle dérivé |
| ΔiHH | Différence d'iHH |
| A | Adénosine |
| ADN | Acide Désoxyribonucléique |
| ALR | *AIM2-like receptors* |
| ARNm | ARN messager |
| ARNt | ARN de transfert |
| C | Cytosine |
| CLR | *C-type lectin receptor* |
| CMH | Complexe majeur d'histocompatibilité |
| CMS | *Composite of multiple signals* |
| CNAS | *Cytosolic nucleic acid sensors* |
| DIND | *Derived Intra-allelic Nucleotide Diversity* |
| DL | Déséquilibre de liaison |
| EHH | *Extended Haplotype Homozygosity* |
| eQTL | *expression Quantitative Trait Locus* |
| G | Guanine |
| GWAS | *Genome-wide association study* |
| iHH | *integrated EHH* |
| iHS | *integrated Haplotype Score* |
| IP | Immunodéficience primaire |
| LRT | *Likelyhood Ratio Test* |
| LSBL | *Locus-Specific Branch Lengt* |
| miARN | microARN |
| MK (test de) | McDonald–Kreitman |
| MTB | *Mycobacterium tuberculosis* |
| MSMD | *Mendelian susceptibility to mycobacterial disease* |
| NK (cellules) | *Natural Killer* |
| NLR | *NOD-like receptors* |
| PAMP | *Pathogen Associated Molecular Pattern* |

| | |
|---|---|
| Pb | Paire(s) de bases |
| PBS | *Population Branch Statistics* |
| pre-miARN | précurseur de miARN |
| pri-miARN | miARN primaire |
| PRR | *Pathogen Recognition Receptors* |
| RISC (complexe) | *RNA-induced silencing complex* |
| RLR | *RIG-I-like receptors* |
| SNP | *Single nucleotide polymorphism* |
| T | Thymine |
| TLR | *TOLL-like receptors* |
| XP-EHH | *cross-Population Extended Haplotype Homozygosity* |

# Préambule

Au cours de son histoire, l'Homme a été confronté à de nombreux micro-organismes qui ont constitué une menace importante pour sa survie et sa procréation. Malgré les progrès médicaux du siècle dernier, les maladies infectieuses restent l'une des principales causes de mortalité à travers le monde. Les êtres humains ne sont pas tous égaux face à ce danger invisible : les individus et les populations présentent d'importantes différences de susceptibilité aux maladies infectieuses. Si certains facteurs environnementaux comme les conditions d'hygiène expliquent partiellement la variabilité de susceptibilité à l'infection, une partie de cette hétérogénéité trouve sa source dans la diversité génétique touchant les gènes impliqués dans la défense de l'hôte ou dans les séquences régulant leur expression. Cette diversité est le fruit de l'histoire démographique et évolutive de notre espèce.

De nombreuses méthodes ont été développées dans le but de comprendre le lien entre diversité génétique et différences de susceptibilité aux maladies infectieuses. Cependant, ces études cliniques, épidémiologiques et évolutives présentent chacune des limites. Alors que les premières ont permis de mettre en évidence des mécanismes indispensables au bon fonctionnement du système immunitaire, elles ne permettent d'expliquer qu'une faible part de la variabilité de réponse face aux infections au niveau de la population. Les études épidémiologiques ont, quant à elles, aidé à identifier des variations génétiques impliquées dans la différence de susceptibilité aux maladies infectieuses mais n'ont mené à l'identification que d'un nombre restreint de mécanismes moléculaires. Finalement, les approches évolutives n'ont porté jusqu'à présent que sur des gènes ou familles de gènes candidats, ne permettant pas d'obtenir une vue globale des forces sélectives ayant participé à l'évolution des gènes de l'immunité.

Au cours de ma thèse, j'ai utilisé une combinaison d'approches *in silico* et *ex vivo* afin d'identifier des variations génétiques participant à la variabilité de réponse immunitaire innée chez l'Homme.

Dans un premier temps, j'ai appliqué un ensemble d'outils de génétique évolutive et de génétique des populations aux données du polymorphisme apportées par le projet 1000 Génomes dans le but d'étudier les signatures moléculaires des pressions sélectives exercées sur les gènes impliqués dans l'immunité innée chez l'Homme. Cependant, si certaines variations génétiques touchent les régions codantes, ils en sont d'autres qui sont localisées dans des séquences régulatrices. Il est donc possible d'identifier des marqueurs génétiques associés à des différences d'expression génique (eQTL). C'est pourquoi j'ai dans un second

temps participé à l'étude de la variabilité de la réponse transcriptionnelle mise en place suite à une stimulation infectieuse. Nous avons utilisé l'expression des miARN au cours de l'infection des cellules dendritiques par *Mycobacterium tuberculosis* comme modèle d'étude et avons combiné un ensemble d'approches *ex vivo* et de techniques génomiques afin de caractériser la régulation de la réponse des cellules à l'infection exercée par les miARN ainsi que le contrôle génétique de l'expression de ces transcrits.

L'objectif de cette thèse est donc d'identifier des facteurs génétiques et des mécanismes moléculaires qui pourraient être impliqués dans les différences de réponse immunitaire innée observées entre les individus dans le but de mieux appréhender les causes de la variabilité de susceptibilité aux maladies infectieuses observée chez l'Homme.

# Introduction

# 1. La diversité génétique

Les caractéristiques physiques et biochimiques des êtres vivants sont variables au sein d'une espèce, ce qui fait que chaque individu est unique. Si certaines dissemblances entre membres d'un même groupe sont expliquées exclusivement par des facteurs environnementaux (chez l'Homme, l'apport insuffisant en vitamine C permet par exemple d'expliquer les cas de scorbut), une grande partie des différences phénotypiques sont liées à des variations au niveau du support de l'information génétique, l'ADN (acide désoxyribonucléique). La diversité génétique de notre espèce résulte de l'action de forces génomiques, démographiques et sélectives. L'étude de ces forces présente un intérêt majeur en génétique des populations et en biologie évolutive car elle permet de retracer l'histoire de notre espèce et de mieux comprendre la distribution de la variabilité génétique et phénotypique des populations. Dans cette partie, je discuterai de l'étendue de la diversité génétique dans l'espèce humaine et présenterai les différentes forces qui la créent et la modèlent.

## 1.1. La diversité génétique de l'espèce humaine

Hormis les jumeaux monozygotes qui partagent la même information génétique, l'ADN de chacun est unique. Il contient l'ensemble des données nécessaires à l'établissement d'un organisme complet et en détermine en grande partie les propriétés. Notre espèce est relativement récente puisqu'on estime que les premiers *Homo sapiens* sont apparus il y a 100 000 à 200 000 ans (Campbell and Tishkoff, 2008; Cavalli-Sforza and Feldman, 2003; Chen et al., 1995; Fagundes et al., 2007; Ingman et al., 2000; Ingman and Gyllensten, 2001; Thomson et al., 2000; Underhill and Kivisild, 2007). Ce temps relativement court à l'échelle de l'évolution n'a pas permis aux groupes d'individus de se différencier de manière importante (Harpending and Rogers, 2000; Li and Sadler, 1991). Deux personnes non apparentées prises au hasard dans la population humaine partagent en moyenne 99,9 % de leur information génétique (Lander et al., 2001; Venter et al., 2001), ce qui est bien plus élevé que chez les autres primates (Fischer et al., 2004; Lander et al., 2001; Thalmann et al., 2007; Venter et al., 2001). Cependant, les 0,1 % de variations permettent d'expliquer de nombreuses différences phénotypiques entre individus. Si certaines sont facilement observables, ils en sont d'autres plus subtiles. L'un des exemples les mieux documentés est celui de la substitution de la cytosine (C) localisée 13 910 paires de bases (pb) en amont du gène codant pour la lactase en thymine (T). Cette mutation permet aux individus qui la portent de digérer le lait à l'âge adulte puisqu'elle engendre un maintien de l'expression du gène codant pour l'enzyme permettant de métaboliser le lactose au-delà de l'enfance (Bersaglieri et al., 2004; Tishkoff et al., 2007; Voight et al., 2006; se référer à Troelsen, 2005 pour une revue des travaux antérieurs).

Seul 1 % des 3,1 milliards de nucléotides dont est composé le génome humain participe à la détermination de la séquence et de la structure des holoprotéides (ENCODE Project Consortium, 2012). La majorité de la diversité génétique présente dans l'espèce humaine touche donc des régions génomiques autres que les exons. De récentes études ont montré que 80 % du génome humain a des propriétés biochimiques particulières qui incluent non seulement les gènes mais aussi des régions non-codantes, des régions régulatrices ainsi que d'autres régions importantes pour la conformation de la chromatine (ENCODE Project Consortium, 2012; Kellis et al., 2014). On comprend dès lors que l'occurrence d'une mutation toutes les 1 000 pb qui est en moyenne observée en comparant deux génomes

humains (ENCODE Project Consortium, 2012; Venter et al., 2001) permet d'expliquer la richesse phénotypique de notre espèce.

Les variations génétiques peuvent prendre de nombreuses formes, allant de la substitution d'un nucléotide par un autre, appelée en anglais *single nucleotide polymorphism* (SNP), à la variation du nombre de répétitions d'un motif donné de plusieurs bases, en passant par l'insertion ou la délétion d'un ou plusieurs nucléotide(s), sans oublier les inversions de motifs nucléotidiques. De manière assez marquante, il a été montré récemment que chaque individu est porteur d'un grand nombre de mutations non-synonymes prédites comme étant fonctionnelles mais n'ayant pas de conséquences phénotypiques pour la plupart. Ainsi, chaque génome humain contient de 340 à 400 mutations touchant 250 à 300 séquences codantes qui impliquent l'arrêt de l'expression de ces gènes ou mènent à la production de protéines non fonctionnelles. De plus, on estime que chaque être humain est hétérozygote pour 50 à 100 sites associés à des maladies mendéliennes, c'est à dire des maladies génétiques monofactorielles héréditaires (1000 Genomes Project Consortium, 2010). De nombreuses études de génétique évolutive et de génétique des populations ont permis de mettre en évidence les forces génomiques, démographiques et sélectives créant, maintenant ou diminuant cette diversité génétique (Figure 1). Ces analyses ont aussi montré que l'occurrence de certaines de ces forces ainsi que l'intensité de leurs actions sont variables entre les individus et les populations (se référer à Jobling et al., 2013 pour une revue).

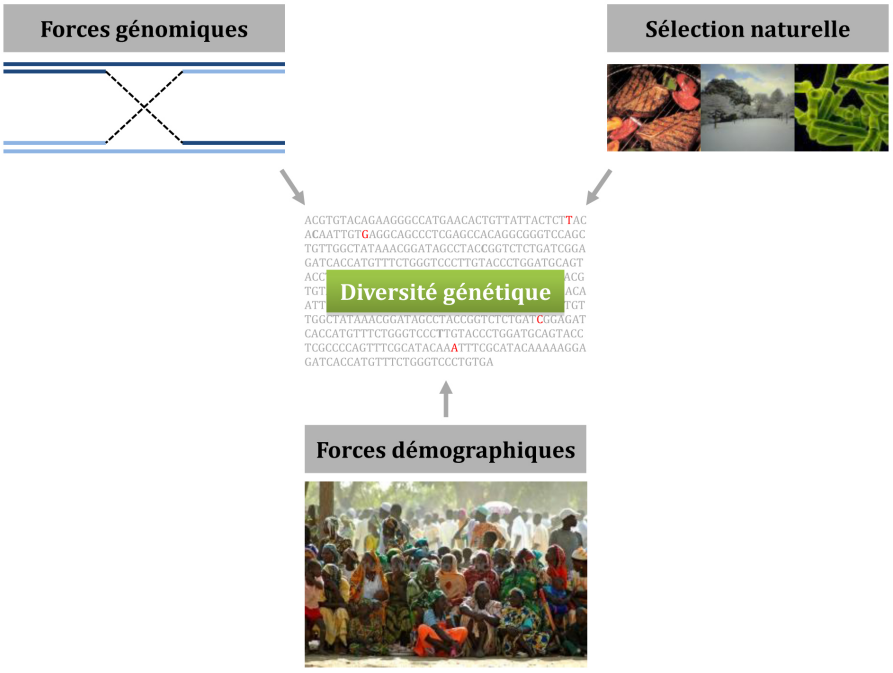



**Figure 1. Les différentes forces à l'origine de la diversité génétique.**

## 1.2. Les forces à l'origine de la diversité génétique

### 1.2.1. Les forces génomiques

Les mutations et les évènements de recombinaison sont les forces qui créent la diversité génétique, respectivement à un site donné ou au sein d'une région génomique. Les variations nucléotidiques les plus abondantes et les plus étudiées en génétique des populations sont les mutations ponctuelles. La version actuelle de dbSNP (Sherry et al., 2001) en recense plus de 97,5 millions validées (build 144, Juin 2015). On peut les classer en deux groupes : les transitions, qui correspondent à la substitution d'une purine par une autre (adénosine (A) <-> guanine (G)) ou d'une pyrimidine par une autre (T <-> C), sont les plus probables (Nei, 1987); les transvertions, elles, consistent en une substitution d'une purine par une pyrimidine, ou inversement. Elles apparaissent suite à des erreurs de réplication ou à l'action de facteurs environnementaux de nature physique tels que les radiations ionisantes ou les rayons ultraviolets, ou de nature chimique tels que les agents mutagènes (Butler and Smith, 1950; Chun et al., 1969; Cross et al., 1987; Kohn et al., 1966; Scholes et al., 1949; Taylor et al., 1947). Elles sont ensuite maintenues par un défaut de la machinerie de réparation de l'ADN (De Bont and van Larebeke, 2004; Friedberg, 2003; Friedberg et al., 2006). Le taux de substitution est estimé à $10^{-8}$ par base et par génération. Il est cependant variable en fonction des régions génomiques étudiées et peut parfois atteindre $10^{-5}$ substitutions par base et par générations (Campbell et al., 2012). Les mutations qui apparaissent dans la lignée germinale sont transmises à la descendance et peuvent devenir polymorphes si leur fréquence dépasse le seuil arbitraire de 1 % dans la population.

Une mutation ponctuelle apparaît dans un contexte génomique particulier. Elle est donc associée à d'autres SNP qui sont à proximité. À cause de leur localisation chromosomique proche, ces allèles vont préférentiellement ségréger ensemble. Ces SNP ne sont donc pas en situation d'indépendance génétique mais sont en déséquilibre de liaison (DL). Le « bloc» qui porte cette combinaison d'allèles est appelé haplotype. Au cours de la méiose, l'appariement des chromosomes homologues peut mener à des réarrangements par recombinaison, ce qui va aboutir au remaniement de certains haplotypes. De nouvelles combinaisons alléliques sont ainsi créées. Tout comme le taux de substitution, le taux de recombinaison est variable le long du génome humain. Il dépend de différents facteurs tels que la distance au centromère, la densité en CG ou la densité en gènes (International HapMap Consortium, 2005; Lichten and Goldman, 1995; McVean et al., 2004; Myers et al., 2008). Ce

brassage allélique est une source de création de diversité génétique s'étendant sur plusieurs pb. Le devenir de cette diversité dépend ensuite de l'action de forces démographiques et de la sélection naturelle.

### 1.2.2. Les forces démographiques

#### 1.2.2.1. La dérive génétique

En l'absence de sélection naturelle et de migration, les fréquences alléliques au sein d'une population varient de manière aléatoire (Nei, 1987; Wright, 1931). Ce phénomène est appelé dérive génétique. Ses effets sont d'autant plus importants que la taille de la population est restreinte. En effet, l'ensemble des allèles présents dans une population à une génération donnée n'est que le fruit d'un échantillonnage aléatoire du réservoir allélique de la génération précédente. Pour que la totalité des allèles d'une génération soit transmis à la suivante, il faut que le nombre de descendants tende vers l'infini. Les fréquences alléliques restent dans ce cas stables au cours des générations. Plus le nombre d'individus pouvant se reproduire dans une population où les rencontres sont aléatoires, appelé taille efficace de la population ou $N_e$, est faible, plus on se s'éloigne de cette situation. L'échantillonnage se fait alors dans un réservoir très restreint, ce qui peut aboutir à d'importantes différences de fréquences alléliques entre deux générations. Ainsi, en suivant ce modèle, une mutation qui vient d'apparaître peut : (i) être éliminée, (ii) augmenter en fréquence jusqu'à fixation (sa fréquence aura atteint 1) ou (iii) être maintenue dans la population au cours du temps. Les effets de la dérive génétique sont particulièrement importants lors de l'apparition d'une espèce, après un goulot d'étranglement, ou dans le cadre d'un isolat génétique. Dans l'ensemble de ces conditions, la diversité génétique sera réduite (1000 Genomes Project Consortium, 2010; Peltonen et al., 1995).

#### 1.2.2.2. Les migrations

La migration correspond au déplacement d'un groupe d'individus de son lieu de vie vers une nouvelle zone géographique. Ce groupe peut créer une nouvelle population ou se mélanger avec une population déjà installée à cet endroit.

Contrairement à la dérive génétique, les migrations n'affectent pas les fréquences alléliques au niveau de l'espèce entière. En revanche, elles peuvent les modifier

indirectement dans une population donnée. Par exemple, les migrations peuvent mener à une réduction de la taille efficace de la population dans laquelle les effets de la dérive génétique seront importants. À l'inverse, si le déplacement d'individus est suivi d'un mélange avec une population préexistante, la diversité génétique de la population receveuse va augmenter. On parle alors de flux génétique ou d'introgression (voir section 1.2.2.3).

Ces dernières années, de nombreuses études ont porté sur la diversité génétique présente sur l'ADN mitochondrial ou le chromosome Y dans le but de reconstruire l'histoire migratoire de notre espèce (Underhill and Kivisild, 2007). Les marqueurs génétiques portés par ces séquences d'ADN n'étant présents qu'en une seule copie dans les cellules de chaque individu, ils s'affranchissent des évènements de recombinaison méiotiques et ne sont soumis qu'à l'apparition de nouvelles mutations. En tenant compte du taux de mutation, il est donc possible de reconstruire la généalogie des haplotypes retrouvés dans le génome mitochondrial et sur le chromosome Y et de dater les nœuds des arbres phylogénétiques ainsi obtenus (Cann et al., 1987; Cavalli-Sforza and Feldman, 2003; Underhill and Kivisild, 2007).

Les données obtenues grâce à cette stratégie montrent que notre espèce est apparue il y a 100 000 - 200 000 ans en Afrique (Cavalli-Sforza and Feldman, 2003; Chen et al., 1995; Ingman et al., 2000; Ingman and Gyllensten, 2001; Underhill and Kivisild, 2007). Elles soutiennent le modèle de « remplacement rapide » qui stipule que les populations non africaines modernes résultent de la sortie côtière d'un ancêtre commun d'Afrique sub-saharienne vers l'Asie et le reste du monde (Figure 2, (Cann et al., 1987; Cavalli-Sforza and Feldman, 2003; Gunnarsdóttir et al., 2011; Jobling and Tyler-Smith, 2003; Macaulay et al., 2005; Quintana-Murci et al., 1999; Thangaraj et al., 2005; Underhill and Kivisild, 2007).

Des données autosomales obtenues récemment (Excoffier et al., 2013; Fagundes et al., 2007; Gravel et al., 2011; Hellenthal et al., 2008; Laval et al., 2010) mettent en évidence la grande diversité génétique observée dans les populations africaines par rapport aux autres populations ainsi que le fait que la diversité présente hors Afrique semble n'être qu'un échantillonnage de celle observée dans ce continent (1000 Genomes Project Consortium, 2010; International HapMap 3 Consortium, 2010; International HapMap Consortium, 2007), corroborant aussi ce modèle.


**Figure 2. Origines de l'Homme moderne et routes migratoires déterminées à partir de données génétiques.**

#### 1.2.2.3. Flux génétique des hommes archaïques à l'homme moderne

Comme nous l'avons mentionné précédemment, les migrations ont joué un rôle majeur dans l'histoire évolutive de l'espèce humaine. Le modèle de la sortie d'Afrique aujourd'hui communément admis suggère qu'au cours de sa colonisation de l'Eurasie et de l'Océanie, l'Homme moderne s'est retrouvé au contact de groupes archaïques qu'il a peu à peu supplantés. Des données archéologiques montrent que l'Homme de Néandertal était présent en Europe et en Asie de l'Ouest il y a 40 000 ans (Higham et al., 2014). La comparaison de cette estimation avec la datation des sites archéologiques associés aux Hommes modernes les plus anciens en Europe suggère qu'*Homo sapiens* et *Homo neanderthalensis* ont pu cohabiter pendant 2 600 ans au moins dans cette région (Higham et al., 2014).

Ces dernières années, les techniques de séquençage d'ADN anciens ont été grandement améliorées et les séquences de l'intégralité des génomes de plusieurs représentants de groupes d'Hommes archaïques sont aujourd'hui disponibles (Castellano et al., 2014; Green et al., 2010; Meyer et al., 2012; Prüfer et al., 2014; Reich et al., 2010). Ces données ont permis d'établir que des croisements entre Hommes modernes non africains et Hommes archaïques se sont faits il y a 50 000 à 60 000 ans (Fu et al., 2014; Sankararaman et al., 2012; Seguin-Orlando et al., 2014). Les Hommes de Néandertal ont ainsi contribué au

génome des populations eurasiennes à hauteur maximale de 4% et les Dénisoviens ont contribué au génome des populations malaisiennes modernes à un taux de 6% (Green et al., 2010; Prüfer et al., 2014; Reich et al., 2010). Le fait que les populations asiatiques présentent une introgression d'allèles archaïques plus importante atteste d'une histoire démographique complexe qui pourrait impliquer un second évènement d'introgression dans ces populations ou d'une dilution des allèles de Néandertal dans les populations européennes due à un mélange avec une population ancestrale non identifiée (Vernot and Akey, 2015). Une étude récente atteste aussi d'un faible taux d'introgression d'allèles Dénisoviens dans les populations natives américaines et celles d'Europe de l'Est (Qin and Stoneking, 2015). Enfin, de récents travaux suggèrent que certaines populations africaines portent elles aussi des traces de mélanges génétiques avec des groupes d'Hommes archaïques encore indéterminés (Hammer et al., 2011; Lachance et al., 2012; Plagnol and Wall, 2006).

Différentes méthodes permettant d'estimer l'intensité du flux de gènes archaïques dans le génome des populations modernes ont été développées ces dernières années (pour une revue, voir Racimo et al., 2015). L'utilisation d'une de ces méthodes, dont les principes sont détaillés en Figure 3, a révélé que l'introgression n'est pas homogène le long des génomes des populations européennes, asiatiques et océaniennes. En effet, la majorité des régions codantes présentent un faible taux d'introgression archaïque (Sankararaman et al., 2014). Cette structure est aussi observée dans les gènes fortement exprimés dans les testicules et le chromosome X ne porte que peu d'allèles introgressés. Ces données suggèrent que les allèles provenant des Hommes archaïques ont été en partie purgés chez les Hommes modernes à cause de la diminution de la fertilité des descendants mâles issus de croisements entre *Homo sapiens* et Hommes de Néandertal (Sankararaman et al., 2014).

À l'inverse, des haplotypes introgressés dans des séquences codantes ont été maintenus chez les Hommes modernes. Par exemple, certains allèles provenant de Néandertal retrouvés dans les génomes des populations non africaines sont liés à des fonctions immunitaires (Sankararaman et al., 2014). Des haplotypes archaïques couvrant des gènes tels que *HLA*, *OAS* et *STAT2* sont trouvés à fortes fréquences dans les populations européennes, asiatiques et malaisiennes (Abi-Rached et al., 2011; Ding et al., 2014b; Mendez et al., 2012a, 2012b; Mendez et al., 2013; Temme et al., 2014), même si ces données sont parfois sujettes à controverse (Ding et al., 2014a).

L'immunité n'est pas la seule fonction biologique ayant bénéficié d'un apport de diversité génétique (voir Racimo et al., 2015 pour une revue exhaustive des travaux actuels) puisque les gènes codant pour la kératine présentent eux aussi un fort taux d'introgression (Sankararaman et al., 2014). De même, le mélange des populations tibétaines avec des populations archaïques semble leur avoir conféré un avantage quant à l'adaptation à l'altitude (Huerta-Sánchez et al., 2014). Ces données attestent donc d'un flux génétique d'homininés aujourd'hui éteints dans le génome de diverses populations actuelles qui a ensuite été soumis à l'action complexe d'une autre force : la sélection naturelle.

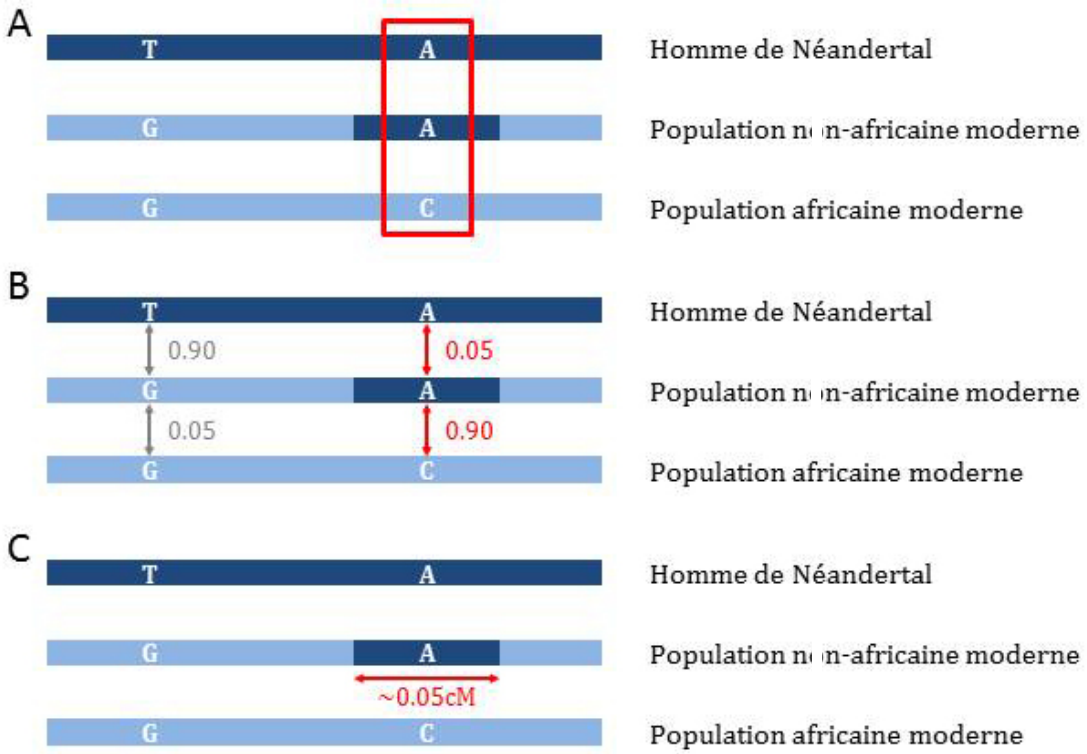



**Figure 3. Propriétés pouvant être testées pour estimer l'ascendance archaïque d'un allèle.**
Adapté de Sankararaman et al. (2014). (A) Pour un SNP donné, si un ensemble d'individus d'Afrique Sub-saharienne portent l'allèle ancestrale et que le Néandertalien séquencé ainsi que l'haplotype retrouvé dans les populations non africaines portent l'allèle dérivé, alors il est probable que celui-ci ait été introgressé de la population archaïque. (B) Si la divergence entre l'haplotype retrouvé dans la population non africaine et les séquences de l'Homme de Néandertal est faible alors que la divergence entre populations africaines et non africaines est élevée, la région a probablement été introgressée des homininés archaïques. (C) Pour qu'un segment soit considéré comme provenant de l'Homme de Néandertal, il doit avoir une taille génétique correspondant à ce qui est attendu sous l'hypothèse d'un mélange entre populations ayant eu lieu il y a 47 000 à 65 000 ans (soit environ 2 000 générations).

### 1.2.3. La sélection naturelle

#### 1.2.3.1. Concept

La notion de sélection naturelle fut formulée par Charles Darwin en 1859 dans son célèbre ouvrage : *On the origin of species by means of natural selection, or, the Preservation of favoured races in the struggle for life* (Darwin, 1859). Dans ce livre, basé sur les observations qu'il a faites au cours de ses voyages, Darwin note que certains groupes d'animaux vivant dans des régions géographiques distinctes partagent de nombreuses caractéristiques. Mais il remarque aussi qu'ils présentent des traits qui leur sont spécifiques et qui semblent avoir participé à leur adaptation à leur environnement. Ainsi, il émet l'hypothèse que ces espèces descendent d'une population unique qui s'est scindée en deux et que chacun des groupes d'individus a par la suite accumulé suffisamment de différences physiques pour qu'ils soient aujourd'hui considérés comme deux espèces distinctes.

Ces variations, apparues spontanément, ne se sont maintenues dans la population que parce qu'elles confèrent un avantage aux individus qui les portent au sein de l'environnement dans lequel ils évoluent. Reprenant les travaux de Thomas Malthus, Charles Darwin formule alors la théorie suivante : comme il naît toujours plus d'individus que le milieu ne peut en nourrir, une lutte pour la vie se met en place entre les êtres vivants de la même espèce ainsi que ceux des espèces vivant dans le même environnement ; seuls arrivent à survivre (et en conséquence procréer) les plus adaptés à leurs conditions de vie, engendrant une élimination des variations défavorables par la sélection naturelle qui ne retient que celles qui sont avantageuses. Ainsi, la sélection naturelle agit sur un trait qui (i) varie dans la population, (ii) est transmissible à la génération suivante et (iii) est lié à un succès reproducteur des individus qui le portent.

On peut regretter que Charles Darwin n'ait pas lu les travaux de son contemporain Gregor Mendel qui, se basant sur des études botaniques, décrivit l'héritabilité de « facteurs », posant les bases de ce que nous appelons aujourd'hui la génétique. Il fallut attendre le XXème siècle pour que naisse la théorie synthétique de l'évolution. Ce concept considère les gènes comme unité de support de l'information génétique et propose une action de la sélection naturelle sur les mutations survenues par hasard dans la population.

#### 1.2.3.2. Les types de sélection naturelle

Une fois le concept formulé, trois grands types de sélection naturelle ont été décrits : la sélection purificatrice (ou négative), la sélection positive (ou darwinienne) et la sélection balancée (Figure 4).

La sélection purificatrice agit sur les mutations qui ont un effet fortement délétère sur l'organisme dans l'environnement dans lequel l'individu évolue. L'idée sous-jacente est que tout porteur d'un variant génétique très défavorable à sa procréation ou à l'atteinte de l'âge de maturité sexuelle a une descendance beaucoup moins nombreuse que celle des non porteurs, voire n'en a pas. La fréquence de la mutation diminue donc rapidement jusqu'à ce que ce variant soit éliminé de la population. Dans le cas d'une mutation défavorable mais non délétère, le variant peut persister à faible fréquence dans la population. Les signatures moléculaires laissées par la sélection purificatrice sont une diminution locale de la diversité génétique et une augmentation du nombre d'allèles rares (Figure 4B, Lohmueller et al., 2011; Nielsen, 2005). On s'attend à ce que ce type de sélection agisse principalement sur les gènes dont la fonction est essentielle pour l'organisme.

La sélection positive agit quant à elle sur les allèles conférant un avantage sélectif aux individus qui les portent. Elle induit une augmentation rapide de la fréquence de ces mutations dans la population et traduit une adaptation à l'environnement dans lequel les individus évoluent. Les conséquences de son action sur un allèle sont une réduction locale de la diversité génétique et l'augmentation du LD, du nombre d'allèles rares et très fréquents et de la différenciation entre populations (Figure 4C, Lohmueller et al., 2011; Nielsen, 2005; Pritchard et al., 2010; Scheinfeldt and Tishkoff, 2013; Vitti et al., 2013).

Finalement, la sélection balancée s'exerce sur un site multiallélique et maintient la coexistence de plusieurs allèles à ce SNP. Plusieurs modèles de sélection balancée ont été proposés. Dans le premier, l'état hétérozygote est plus avantageux que les deux états homozygotes. On parle alors de superdominance. Le second modèle repose sur l'avantage octroyé par un phénotype en fonction de sa fréquence relativement aux autres phénotypes. On parle alors de sélection dépendante de la fréquence. Cette sélection peut être positive ou négative, selon que l'avantage sélectif apporté par le phénotype augmente ou diminue avec sa fréquence dans la population. Le dernier type de sélection balancée consiste en une oscillation du génotype le plus avantageux dans le temps (au cours de la vie de l'individu) ou dans l'espace (en fonction de l'environnement). La sélection balancée est le seul type de

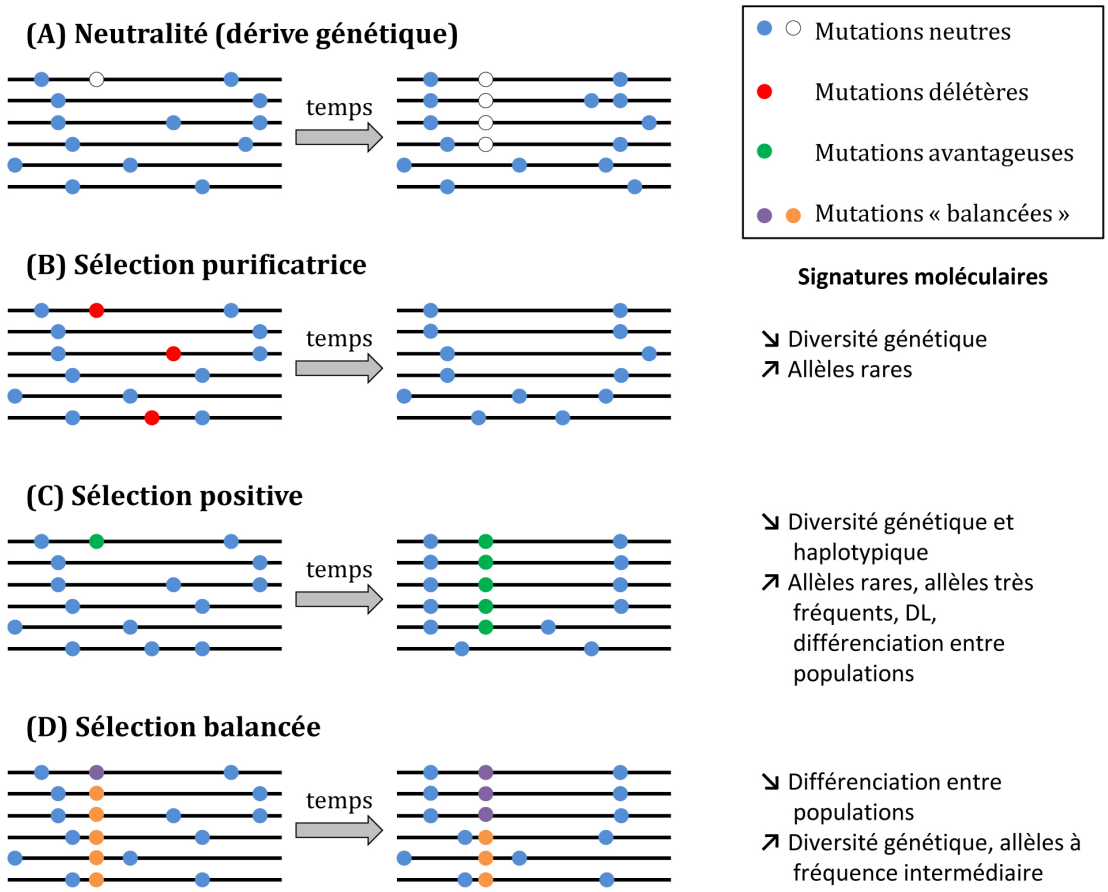



**Figure 4. Les différents types de sélection naturelle et leurs signatures moléculaires.** Évolution d'une région génomique sous (A) neutralité, (B) sélection purificatrice, (C) sélection positive et (D) sélection balancée. Les points représentent les mutations, la couleur bleue et la couleur blanche indiquant un effet neutre pour l'organisme et les autres couleurs indiquant un effet de la sélection sur la variation génétique.

sélection naturelle qui engendre une augmentation de la diversité génétique et des fréquences alléliques intermédiaires. Elle induit aussi une diminution de la différenciation entre populations (Figure 3D, Charlesworth, 2006; Nielsen, 2005).

Chacun de ces types de sélection naturelle laisse des signatures moléculaires spécifiques dans le génome qu'il est possible d'étudier afin de retracer l'histoire évolutive et adaptative des populations.

#### 1.2.3.3. Détection de la sélection naturelle

Détecter les régions du génome humain présentant des signatures moléculaires de sélection naturelle permet de reconstruire l'histoire évolutive de l'Homme et de comprendre à quels facteurs environnementaux notre espèce s'est adaptée. Cela permet aussi d'inférer

l'importance fonctionnelle de certaines régions génomiques. Afin d'identifier les traces moléculaires laissées par la sélection naturelle sur le génome humain, il faut dans un premier temps estimer l'évolution des mutations en l'absence de sélection. En 1968, Kimura formula la théorie neutre de l'évolution moléculaire. À travers cette théorie, il considère que la vaste majorité des variations génétiques entre et au sein de chaque espèce n'a pas d'effet sur la survie ou le succès reproductif des individus qui les portent. Les mutations ont donc un effet neutre sur l'organisme et évoluent par dérive génétique, c'est à dire que la variation des fréquences alléliques au cours des générations est stochastique (Kimura, 1968a, 1968b). Kimura considère aussi les mutations qui ont des effets sur l'organisme mais estime que celles qui sont fortement délétères sont rapidement éliminées de la population et n'ont que peu d'effets sur les données de polymorphisme et de divergence. Quant aux mutations avantageuses, il estime qu'elles sont très peu nombreuses et qu'elles arrivent rapidement à fixation dans la population.

La théorie neutre de l'évolution est basée sur un modèle mathématique supposant que : (i) la taille de la population reste constante, (ii) le taux de mutation est constant dans le temps et (iii) il y a un état d'équilibre entre le nombre d'allèles perdus par dérive génétique et le nombre de nouveaux allèles créés par mutation. Cette théorie est utilisée comme hypothèse nulle dans les études visant à étudier l'action de la sélection naturelle sur notre génome. Ainsi, plus les pressions de sélection sont importantes, plus les observations faites sont éloignées de celles modélisées par la théorie neutre de l'évolution. De nombreux tests visant à identifier les empreintes moléculaires laissées par la sélection sur le génome grâce à la mesure des écarts à la neutralité ont été développés. Ils peuvent se diviser en deux groupes : les tests interspécifiques, qui visent à comparer des espèces différentes, et les tests intraspécifiques, qui s'intéressent aux différences observées entre les populations d'une espèce (se référer à Vitti et al. (2013) pour une revue des statistiques qu'il est possible d'utiliser). Nous évoquerons ici un certain nombre de tests et discuterons de ceux qui ont été utilisés au cours de cette thèse.

*Tests de neutralité interspécifique et sélection purificatrice*

Les tests interspécifiques reposent sur des principes de génomique comparative. Ils testent pour un excès ou un défaut de divergence entre deux espèces génétiquement proches. Ils permettent de mettre en évidence des séquences d'ADN correspondant à des régions codantes

ou non-codantes dont l'importance biologique n'autorise pas ou peu de variation génétique altérant leur fonction. Les séquences non fonctionnelles, elles, sont libres de changer. Les tests interspécifiques ont été rendus possibles grâce au développement des techniques de séquençage et à leur application à de nombreuses espèces dont l'Homme, pour lequel les données génétiques d'un grand nombre d'individus sont disponibles. Ils sont utilisés pour mettre en évidence des traces de sélection très anciennes. Par exemple, lors d'une comparaison du génome humain avec celui du chimpanzé, les différences étudiées se sont accumulées pendant les 6 millions d'années au cours desquelles les deux espèces ont évolué séparément. L'ensemble de la variabilité génétique de chaque espèce est considéré, ce qui fait que les tests interspécifiques ne permettent pas d'identifier des évènements sélectifs ayant eu lieu dans une population spécifique.

En 1991 fut développé le test de McDonald-Kreitman (MK) qui prend non seulement en compte les données de divergence entre deux espèces mais considère aussi le niveau de polymorphisme des espèces considérées (McDonald and Kreitman, 1991). Ce test émet l'hypothèse que les mutations synonymes sont neutres et ne sont pas soumises à la sélection naturelle, contrairement aux mutations non-synonymes qui sont fonctionnelles et sujettes à l'action de la sélection. Le test de MK établit ainsi une table de contingence puis évalue la non-indépendance entre les variations génétiques non-synonymes (divergence et polymorphisme) et les variations neutres à l'aide d'un test de Fisher.

Il est important de noter que le test de MK implique un certain nombre d'approximations pouvant créer un biais au niveau des résultats. En effet, toute mutation synonyme y est considérée comme neutre, non soumise à la sélection naturelle. Nous l'avons évoqué précédemment, des propriétés fonctionnelles ont été attribuées à 80 % des séquences du génome humain (ENCODE Project Consortium, 2012). Il est donc possible qu'une mutation silencieuse soit localisée dans une de ces régions. On peut aussi envisager qu'une variation génétique induise un écart au biais de l'usage du code génétique qui, sous l'hypothèse de sélection traductionnelle, suggère que certains codons codant pour un acide aminé sont plus utilisés que d'autres car ils correspondent à l'ARN de transfert (ARNt) le plus abondant dans la cellule (Ikemura, 1981) et permettent une plus grande rapidité ou une plus grande fiabilité de la traduction de l'ARN messager (ARNm). Ces mutations fonctionnelles ne sont pas considérées par le test de MK mais peuvent être soumises à la sélection. À l'inverse, certaines modifications de la séquence d'acides aminés d'une protéine

sont tolérées car peu dommageables pour la structure ou la fonction de la protéine. La classification des mutations réalisée dans le cadre du test MK ne prend pas en compte ces subtilités. En revanche, l'un des avantages de cette analyse est qu'elle n'est que peu sensible à l'histoire démographique des populations, celle-ci impactant de manière égale les mutations synonymes et non-synonymes.

Les tests interspécifiques réalisés au cours de cette thèse ont été faits à l'aide de la méthode SnIPRE (Eilertson et al., 2012) qui se base sur les données de polymorphisme et de divergence utilisées dans le cadre du test MK pour identifier les gènes sous sélection. Cette méthode intègre un modèle linéaire généralisé mixte pour considérer des données pangénomiques comme effets fixes et des données sur les gènes pris individuellement comme effets aléatoires. Cela permet de grouper les informations de plusieurs gènes et d'augmenter le pouvoir de détection des séquences codantes sous sélection. Le paramètre *f* de cette méthode estime la proportion de mutations non-synonymes délétères dans les gènes. Si l'ensemble des mutations délétères a été purgé de la population humaine, ce paramètre prend une valeur nulle. Sa valeur sera de 1 si, au contraire, l'ensemble des mutations non-synonymes a un effet neutre sur l'organisme. Nous avons donc utilisé ce paramètre dans le but de quantifier l'action de la sélection purificatrice sur les gènes humains.

*Tests intra-spécifiques et sélection positive*

- Méthodes basées sur les fréquences alléliques.

En 1983, Tajima détermina la distribution des fréquences alléliques dans une population sous hypothèse de neutralité (Tajima, 1983). Plusieurs tests ont ensuite été développés dans le but de tester si le spectre de fréquences alléliques observé dans une région donnée correspond à celui attendu sous neutralité. Nous pouvons notamment citer le *D* de Tajima (Tajima, 1989), le *D* et le *F* de Fu et Li (Fu and Li, 1993), le *H* de Fay et Wu (Fay and Wu, 2000)et le *E* de Zeng (Zeng et al., 2006). Ces estimateurs prennent une valeur nulle si le spectre des fréquences alléliques de la région testée est similaire à celui attendu sous hypothèse de neutralité. Sous sélection positive, on observe un enrichissement en allèles à faibles et fortes fréquences ainsi qu'un défaut de fréquences alléliques intermédiaires. Cela se traduit par des valeurs négatives des indicateurs de Fu et Li, Tajima et Zeng. Cependant, la sélection purificatrice et l'expansion de la population mènent à un résultat similaire (l'impact de la démographie sur la détection de la sélection fait l'objet d'une partie page 25). Afin de

discriminer entre sélection négative et positive, il est possible d'utiliser l'estimateur *H* de Fay et Wu qui prend des valeurs négatives en cas de sélection darwinienne.

Il est aussi possible de comparer les fréquences alléliques de l'allèle dérivé à un site donné entre populations (*ΔDAF* pour *Derived Allele Frequency*, Grossman et al., 2010). Cette méthode est plus sensible pour distinguer les allèles sélectionnés que la simple fréquence de l'allèle dérivé.

- Variations de la longueur d'haplotypes.

Une mutation apparaît dans un contexte génomique donné. Elle est en DL avec d'autres allèles présents sur le même chromosome, définissant ainsi un nouvel haplotype. Sous hypothèse de neutralité, une mutation n'arrive à une fréquence élevée qu'après un certain temps au cours duquel la combinaison va augmenter la diversité locale et réduire la taille du bloc de DL comme nous l'avons décrit précédemment. Une mutation sous sélection positive augmente rapidement en fréquence, laissant moins de temps à la recombinaison pour agir (Sabeti et al., 2002; Voight et al., 2006). Le bloc de DL est donc de taille plus importante qu'attendue sous hypothèse de neutralité et la diversité locale plus faible. Plusieurs tests se basent sur ces propriétés pour détecter les régions génomiques sous sélection positive. C'est par exemple le cas de l'*EHH* (*Extended Haplotype Homozygosity*, Figure 5, Sabeti et al., 2002) qui mesure la diminution de l'homozygotie moyenne entre haplotypes en fonction de la distance croissante autour de la mutation d'intérêt. L'*iHH* (Voight et al., 2006), qui est une mesure de l'aire située sous la courbe de l'*EHH*, peut être comparé entre populations ( *Δ iHH*, Grossman et al., 2010). Il peut aussi être utilisé pour calculer l'*iHS* (*integrated Haplotype Score*, Voight et al., 2006) qui s'écarte significativement d'une distribution normale en cas de sélection positive de l'allèle étudié. Ce dernier test a par la suite mené au développement de l'*XP-EHH* (*cross-Population Extended Haplotype Homozygosity*, Sabeti et al., 2007) qui mesure et compare l'*iHS* dans différentes populations.

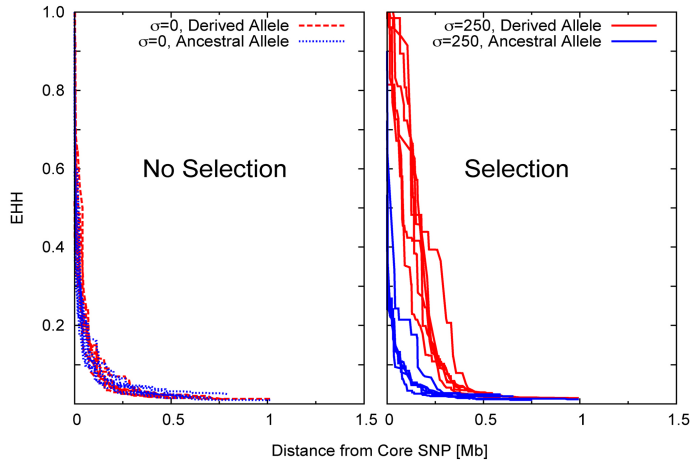



**Figure 5. Effet de la sélection positive ciblant l'allèle dérivé sur profil de diminution de l'*EHH*.**
Tiré de Voight et al. (2006). La première figure représente les courbes d'*EHH* attendues sous neutralité alors que la seconde représente celles observées sous sélection positive de l'allèle dérivé. Sous sélection, l'*EHH* de l'allèle dérivé (en rouge) diminue plus lentement en fonction de la distance au SNP que celui de l'allèle ancestral (en bleu).

Plus récemment, le test *DIND* (*Derived Intra-allelic Nucleotide Diversity*, Barreiro et al., 2009) a été développé pour mettre à profit les données de séquençage. Ce test compare la diversité nucléotidique portée par l'allèle ancestral et celle portée par l'allèle dérivé, en fonction de la fréquence de l'allèle dérivé. Une valeur élevée de *DIND* traduit l'action de la sélection positive sur l'allèle dérivé.

- Différenciation entre populations

Une mutation peut conférer un avantage sélectif à un groupe d'individus évoluant dans un environnement particulier. Elle sera alors sous sélection positive dans cette population et augmentera en fréquence rapidement. En revanche, elle aura un effet neutre dans d'autres populations qui ne sont pas soumises aux mêmes conditions environnementales. Ainsi, à la localisation de cette mutation, la différenciation entre populations augmentera. La statistique $F_{ST}$ permet de mesurer cette différenciation et prend des valeurs comprises entre 0 (pas de différenciation) à 1 (différenciation totale). Sous neutralité, le $F_{ST}$ est déterminé par la dérive génétique. Un évènement de sélection positive locale aura pour effet d'augmenter le $F_{ST}$ au locus considéré, permettant d'identifier directement la variant sous sélection (Cavalli-Sforza, 1966; Lewontin and Krakauer, 1973; Weir and Hill, 2002; Wright, 1965, 1943). En revanche,

cette statistique ne permet pas de déterminer la population dans laquelle la différenciation s'est faite.

D'autres tests se basant sur le $F_{ST}$ mais contournant cette limitation ont été récemment développés. Le *LSBL* (*Locus-Specific Branch Length*, Shriver et al., 2004), le *PBS* (*Population Branch Statistics*, Zhang et al., 2005) et le *LRT* (*Likelyhood Ratio Test*, Bhatia et al., 2011) permettent de reconstruire les arbres génétiques à un locus donné en incluant au moins trois populations. Une comparaison des longueurs des branches deux à deux permet ensuite de déterminer dans quelle population l'évènement adaptatif a eu lieu.

- Développement de méthodes composites

Chacune des méthodes décrites précédemment teste un aspect particulier des signatures moléculaires laissées par la sélection positive dans notre génome. Pour tester différentes propriétés, il est possible d'effectuer plusieurs tests séparément. Cependant, deux problèmes majeurs se posent : (i) chaque analyse présente un certain taux de faux positifs, c'est à dire de sites déterminés par la statistique comme étant sujet à l'action de la sélection positive alors qu'il ne le sont pas en réalité et (ii) la sensibilité des tests fait qu'il est possible de ne pas considérer un site présentant des signatures moléculaires à la limite du seuil de détection de la statistique utilisée (faux négatifs). La balance entre sensibilité et limite du taux de faux positifs est difficile à trouver et utiliser en parallèle plusieurs tests reviendrait à cumuler les erreurs de chacun de ces tests.

Ces dernières années, des méthodes composites ont été développées dans le but d'affiner les résultats des analyses de détection des signaux de sélection positive dans le génome. Deux stratégies ont été utilisées. La première repose sur l'idée qu'un faux signal peut apparaître ponctuellement de façon aléatoire alors qu'un ensemble de signaux localisés dans une même région est plus probablement dû à une véritable signature de sélection (Carlson et al., 2005). Ainsi, l'utilisation d'une même statistique sur l'ensemble des SNP d'une région permet d'augmenter le pouvoir tout en réduisant le taux de faux positifs. C'est sur ce principe que reposent le *CLR* (pour *Composite Likelihood Ratio*, Nielsen et al., 2005b) et son équivalent permettant de comparer les fréquences alléliques dans une région génomique entre populations (*XP-CLR*, Chen et al., 2010). Des approches normalisant les signaux d'*iHS* et de *DIND* sur des fenêtres de taille physique déterminée ont aussi été

utilisées pour mettre en évidence des régions génomiques sous sélection positive (Fagny et al., 2014; Voight et al., 2006).

Une autre stratégie consiste à cumuler les résultats de plusieurs statistiques pour un même site. Le but est ici d'utiliser des approches complémentaires pour obtenir une meilleure résolution spatiale du signal de sélection (Zeng et al., 2006). Une combinaison du *D* de Tajima et du *H* de Fay et Wu permet par exemple de détecter des évènements de sélection positive plus efficacement qu'en utilisant les deux statistiques séparément (Zeng et al., 2006). C'est en se basant sur cette idée que Grossman *et al.* ont développé le *CMS* (*Composite Multiple Score*) qui multiplie les valeurs *P* de plusieurs statistiques testant la longueur des haplotypes (*XP-EHH*, $\Delta$*iHH* et *iHS*) et la différenciation génétique entre populations ($F_{ST}$ et $\Delta$*DAF*) pour ne plus obtenir qu'un score unique pour chaque site testé (Figure 6, Grossman et al., 2013, 2010). Les valeurs *P* utilisées dans ce test ont été obtenues à partir de simulations utilisant un modèle démographique des populations sous hypothèse de neutralité ainsi qu'un évènement de sélection positive (Schaffner et al., 2005). D'autres études ont proposé de résumer les résultats de statistiques permettant de détecter l'action de la sélection positive sur les séquences en suivant la méthode de Fisher (Luisi et al., 2015, 2012; Zaykin et al., 2007).

*Impact de la démographie sur la détection de la sélection*

La démographie a un impact important sur la diversité génétique. Par exemple, une réduction de la taille d'une population correspond à un échantillonnage, ce qui se traduit par une réduction de la diversité génétique de la population étudiée. À l'inverse, une population en expansion voit sa diversité génétique augmenter. Dans ce cas précis, la population présentera un excès d'allèles rares qui est aussi, comme nous l'avons évoqué précédemment, l'une des signatures moléculaires de la sélection positive. Les tests de neutralité utilisés en génétique des populations considèrent dans l'hypothèse nulle une taille de population constante et une absence de structure des populations. Ils se révèlent donc particulièrement sensibles à l'histoire démographique des populations.

La principale différence entre démographie et sélection naturelle réside dans le fait que la première agit indistinctement sur l'ensemble du génome alors que la seconde cible une ou plusieurs régions spécifiques. Pour détecter la sélection positive dans une population donnée, on compare donc les valeurs obtenues sur une région candidate à une distribution

nulle obtenue (i) soit en ne considérant que des régions supposées neutres qui n'ont évolué qu'en fonction de l'histoire démographique de la population, (ii) soit en simulant une histoire démographique réaliste de la population considérée.

Nous avons mentionné dans la partie 1.2.2.2 que l'utilisation de données génétiques avait permis de redessiner les voies migratoires empruntées par *Homo sapiens* lors de sa colonisation du globe. Ces données, combinées à des résultats de simulation, ont permis d'établir un modèle démographique consensus des populations humaines d'Afrique de

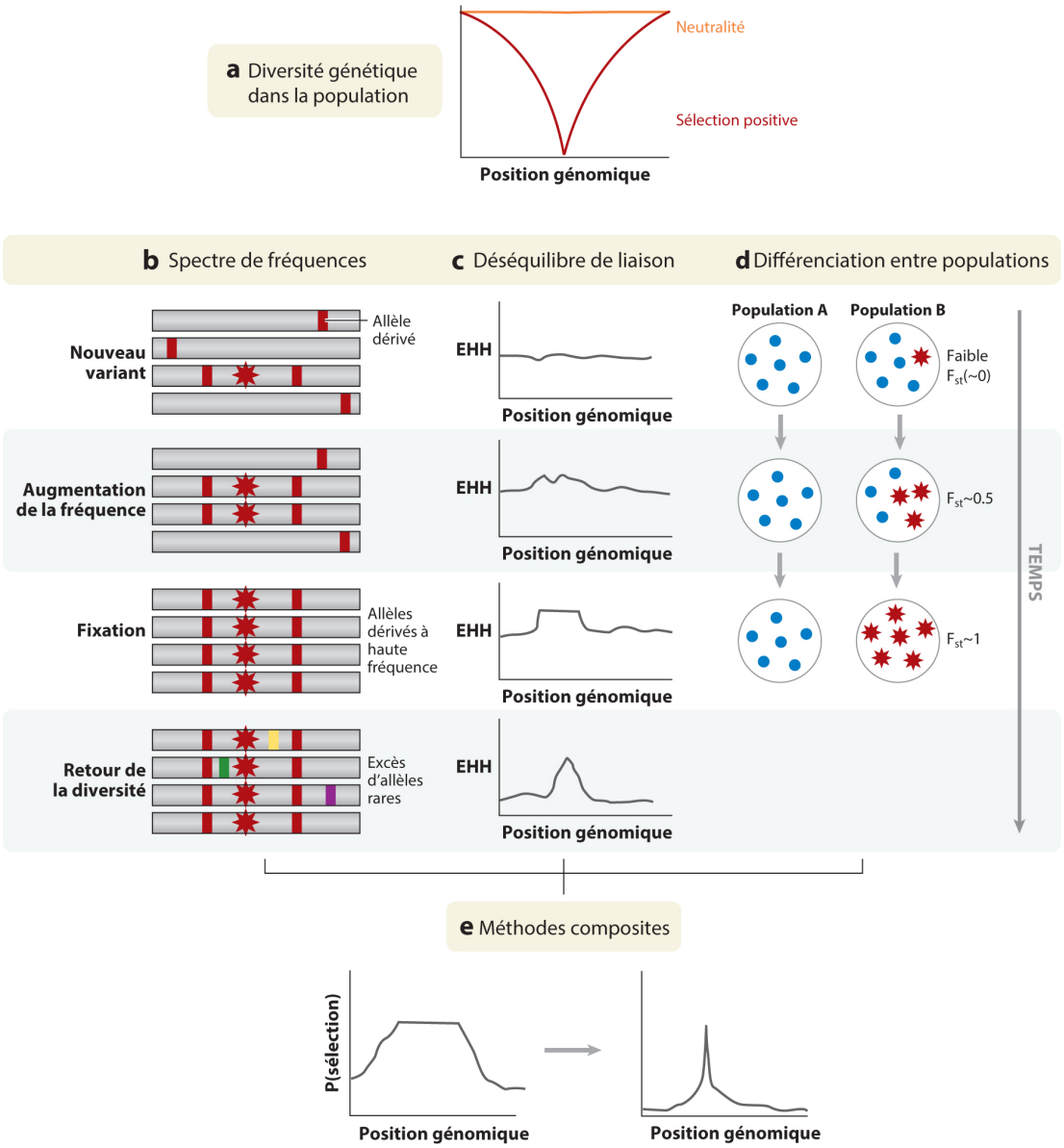



**Figure 6. Stratégie de méthode composite permettant d'affiner la détection de l'allèle ciblé par le balayage sélectif.**

Adapté de Vitti et al. (2013). (a) Lorsqu'elle apparaît, la mutation conférant un avantage sélectif à l'individu qui la porte est en DL avec d'autres allèles. Ceux-ci sont entraînés par le balayage sélectif agissant sur la mutation avantageuse. La diversité génétique locale est donc réduite par rapport à ce qui est attendu sous neutralité. (b) La mutation bénéfique est rapidement fixée. Les allèles dérivés qui sont en DL avec elle ont été entraînés et se retrouvent à des fréquences élevées dans la population. De nouvelles mutations vont survenir dans ce fond génétique homogène, induisant un surplus d'allèles rares. (c) Au début du balayage sélectif, l'haplotype portant la mutation avantageuse présente un défaut de diversité génétique ce qui se traduit par une valeur d'*EHH* élevée. Au cours du temps, l'action des forces génomiques et l'apparition de nouvelles mutations vont créer et remanier la diversité génétique qui et la région d'homozygotie diminuera progressivement autour de la mutation sélectionnée. (d) L'augmentation de la fréquence allélique induit une différenciation entre les populations qui se traduit par des valeurs élevées de $F_{ST}$. (e) Les méthodes composites intègrent l'ensemble de ces signaux afin de faciliter l'identification de la mutation ciblée par le balayage sélectif.

l'Ouest, d'Europe et d'Asie de l'Est (Excoffier et al., 2013; Fagundes et al., 2007; Gravel et al., 2011; Laval et al., 2010; Schaffner et al., 2005). Brièvement, le modèle assume que l'exode hors d'Afrique des populations eurasiennes s'est fait il y a 50 000 à 75 000 ans et s'est accompagné d'un goulot d'étranglement d'une intensité de l'ordre d'environ 5 fois la taille efficace de population (Laval et al., 2010). Les populations européennes et asiatiques ont ensuite divergé il y a 20 000 à 40 000 ans, subissant une nouvelle réduction de la taille efficace de la population. Cette période correspond au début d'une expansion modérée des populations africaines à un taux de 0,007 individus par générations (Laval et al., 2010). L'ensemble des populations a ensuite connu une expansion forte lors de l'avènement de l'agriculture il y a 5 000 à 10 000 ans (Schaffner et al., 2005). L'ensemble de ces données montre que l'histoire démographique de l'espèce humaine est rythmée d'isolements et de migrations.

## 1.3. Le Projet 1000 Génomes, une base de données du polymorphisme humain

Le Projet 1000 Génomes (1000 Genomes Project Consortium, 2010) a pour but de fournir un catalogue particulièrement détaillé des variations génétiques présentes dans la population humaine grâce au séquençage de l'intégralité du génome d'individus appartenant aux principales populations d'Afrique de l'Ouest, d'Europe, d'Asie du Sud Est et des Amériques. Se basant sur les techniques de séquençage à très haut débit, cette collaboration internationale vise à identifier toutes les formes de variations génétiques présentes dans le génome humain à une fréquence supérieure à 1 % pour mieux comprendre la contribution génétique aux phénotypes humains. Afin d'accéder à des variations plus rares et d'identifier des allèles aux conséquences délétères pour l'organisme, le séquençage des exons des individus recrutés pour cette étude est réalisé en parallèle à une profondeur de 50 à 100X. La phase I du projet, qui porte sur 1 092 individus de 14 populations différentes, a ainsi permis de mettre en évidence ~41 millions de SNP, 1.4 millions d'insertions et de délétions d'un faible nombre de nucléotides ainsi que plus de 14 000 délétions de plus grandes tailles (1000 Genomes Project Consortium, 2012). Une approche *in silico* a aussi été conduite dans le but de reconstruire les haplotypes ségrégant dans les populations étudiées et d'inférer l'état allélique à certains locus qui n'ont pas été génotypés. Le consortium a récemment commencé à communiquer sur les

résultats de la phase 3 du projet qui, à terme, sera mené sur ~2500 individus répartis en 27 populations. Ces données permettront d'avoir une vue particulièrement détaillée de la diversité génétique de notre espèce, tant au niveau individuel que populationnel.

# 2. Diversité génétique et maladies infectieuses

Comme nous l'avons précédemment évoqué, une partie de la diversité phénotypique observée au sein de la population humaine est expliquée par la diversité génétique de notre espèce. Si certains phénotypes sont visibles à l'œil nu, ils en sont d'autres qui sont beaucoup plus subtils à étudier. Lors d'une infection par un pathogène, la réponse immunitaire de l'hôte est mise à contribution. Cependant, elle ne permet pas toujours de nettoyer l'organisme des agents infectieux avant l'apparition de symptômes qui varieront en intensité d'une personne à l'autre. Dans cette partie, je présenterai l'établissement du lien entre diversité génétique et maladies infectieuses ainsi que certaines des méthodes mises en œuvre pour identifier les facteurs génétiques de l'hôte impliqués dans la réponse à l'infection.

## 2.1. Concept

Jusqu'au milieu du XIXème siècle, l'espérance de vie à la naissance n'excédait pas les 20-25 ans dans la population humaine (Cairns, 1998). Ceci s'expliquait par les nombreuses fièvres touchant les enfants en bas âge. Les découvertes de certains micro-organismes comme agents contagieux et infectieux au cours de la deuxième moitié du XIXème siècle (Koch, 1882; Pasteur and Gauthier-Villars, 1870) ont ouvert la voie à ce que l'on appelle aujourd'hui l'immunologie. Dès lors que les bases de la théorie microbienne furent posées, observation fut faite que seule une partie des personnes infectées par un même pathogène et présentant de la fièvre succombaient à l'infection. De manière plus subtile, seule une fraction des personnes infectées par un même pathogène développent des symptômes de la maladie infectieuse. Cela suggère que des facteurs inhérents à l'hôte jouent un rôle dans le devenir de l'infection ainsi que dans la susceptibilité aux maladies infectieuses et que ces facteurs sont variables entre les individus.

L'une des premières mises en évidence formelles de ce phénomène fut réalisée en 1949 par John B.S. Haldane qui émit l'hypothèse que des facteurs génétiques présents dans les populations vivant dans des régions endémiques de la malaria pourraient conférer une plus grande protection contre cette maladie infectieuse (Haldane, 1949). Il fut ainsi le premier à faire le lien entre la thalassémie et la protection contre la malaria. Les mécanismes moléculaires et cellulaires sous-jacents furent mis en évidence cinq ans plus tard (Allison, 1954). Ces travaux ouvrirent la voie aux études de génétique moléculaire appliquées au domaine de l'immunologie.

De nombreuses analyses ont par la suite permis une meilleure compréhension du fonctionnement de notre système immunitaire, identifiant des facteurs génétiques impliqués dans la variabilité interindividuelle de réponse immunitaire et de susceptibilité aux maladies infectieuses et auto-immunes. Il est possible de diviser ces études menées sur l'Homme en trois grands types d'approches : les études de génétique clinique, les études d'épidémiologie génétique et les études de génétique évolutive. Nous évoquerons brièvement les deux premiers types d'analyses dans ce chapitre, le dernier faisant l'objet d'une section séparée (voir chapitre 3 de l'introduction).

## 2.2. Les études de génétique clinique

Les études de génétique clinique reposent sur l'étude des Immunodéficiences Primaires (IP) qui sont les cas les plus extrêmes de prédisposition aux maladies infectieuses. Ces maladies, dites mendéliennes, sont monogéniques, rares et résultent de mutations délétères ségrégant dans les familles avec une forte pénétrance (Pritchard and Cox, 2002; Zwick et al., 2000). La fréquence allélique des variations à la base des maladies mendéliennes sévères est habituellement très inférieure à 1 % et est d'autant plus faible que le trouble est sévère (Figure 7, Pritchard and Cox, 2002). Dans la plupart des cas, un faible nombre de patients présentent les mêmes symptômes mais il est parfois possible de n'avoir qu'un unique patient présentant des signes cliniques qui lui sont propres. À ce jour, plus de 250 IP ont été recensées mais il a récemment été suggéré que ces pathologies sont certainement plus nombreuses que les estimations actuelles (Al-Herz et al., 2014; Bousfiha et al., 2013).

La plupart des immunodéficiences monogéniques ont été identifiées par des approches gènes-candidats qui se basent sur les fonctions immunologiques de diverses séquences codant pour des protéines précédemment décrites chez l'Homme ou dans des

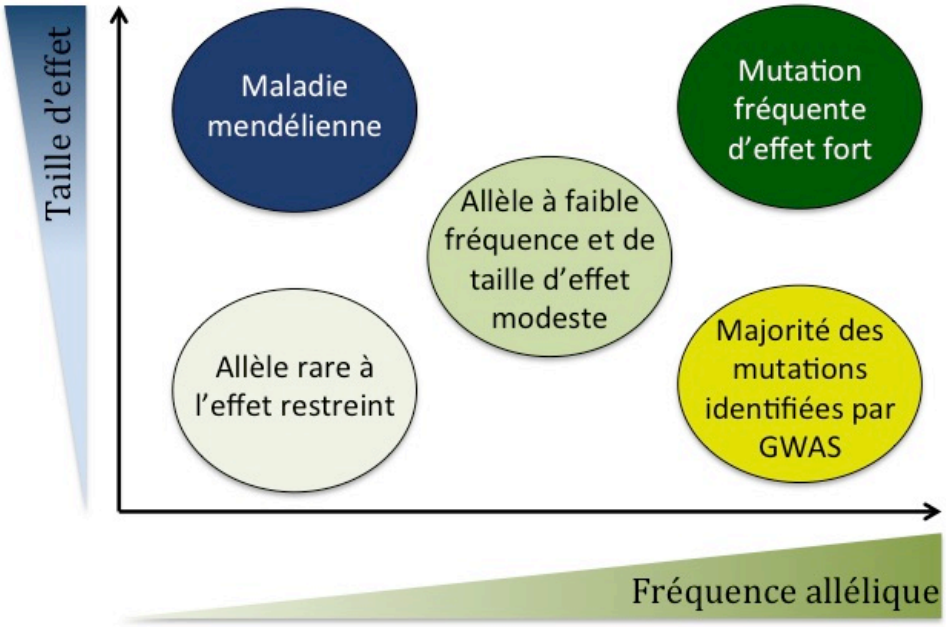



**Figure 7. Représentation schématique de la relation entre les fréquences alléliques des variations génétiques et la force de leur effet sur le phénotype.**
Adapté de McCarthy et al. (2008). La vision dichotomique de variations rares ayant des effets importants sur un phénotype mendélien et de maladies complexes causées par un ensemble de variations génétiques ayant chacune un effet faible sur le trait observé est aujourd'hui plus nuancée. Les phénotypes complexes semblent résulter de l'expression de mutations variant à la fois en fréquence et en taille d'effet.

modèles animaux. Ce type d'approche, couplé à des analyses de liaison portant sur l'ensemble du génome, a par exemple permis de mettre en évidence l'implication d'un défaut de la protéine UNC-93B chez des patients présentant une encéphalite causée par le virus HSV-1 (Casrouge et al., 2006). Cette étude a de particulier qu'elle se penche sur les cas de deux patients présentant une IP caractérisée par une susceptibilité à un nombre restreint de pathogènes, alors que la plupart des IP confèrent une susceptibilité à de multiples infections (J. L. Casanova and Abel, 2004; Chapman and Hill, 2012).

Plus récemment, le développement des techniques de séquençage et notamment de séquençage des exomes a permis l'identification de variations génétiques rares impliquées dans des susceptibilités mendéliennes aux maladies infectieuses et des troubles immunologiques. La première étude utilisant ce type d'approche a mis en lumière l'implication d'une mutation d'un site d'épissage de gène *STIM* dans la maladie de Kaposi qui, bien qu'elle ne soit pas infectieuse, est due à une infection par l'herpèsvirus humain type 8 (Byun et al., 2010).

Les approches utilisées en génétique clinique ont ainsi permis d'identifier des mutations touchant de nombreux gènes responsables d'IP, révélant leur rôle crucial dans la réponse immunitaire (Angulo et al., 2013; Bogunovic et al., 2012; Boisson-Dupuis et al., 2012; Burns et al., 2014; Bustamante et al., 2014; Casanova et al., 2013; Conley and Casanova, 2014; Okada et al., 2015; Pannicke et al., 2013; Pérez de Diego et al., 2010).

## 2.3. Les études d'épidémiologie génétique

Les approches d'épidémiologie génétique combinent des données génétiques et épidémiologiques obtenues à partir de l'étude d'une population. Elles peuvent aussi s'intéresser à l'agrégation familiale d'un trait phénotypique (Khoury et al., 1993). Les avancées technologiques permettant d'identifier et de suivre la ségrégation de marqueurs génétiques ont été critiques pour le développement de ce type d'études. Cependant, la mise en place de ces stratégies n'est pas récent puisque Haldane suivit ce mode opératoire pour établir le lien entre la thalassémie et la protection contre la malaria, comme nous l'avons précisé un petit peu plus tôt (Haldane, 1949).

D'autres études d'épidémiologie génétique ont ensuite mené à l'association de mutations dans les gènes *CCR5* et *DARC* à une protection contre des maladies infectieuses comme le VIH et la malaria, respectivement (Dean et al., 1996; Miller et al., 1976). Outre l'utilisation dans le contexte de l'étude de traits mendéliens, les études d'épidémiologie génétique sont particulièrement puissantes pour identifier la part génétique des phénotypes complexes. Ces traits phénotypiques n'ont pas une transmission mendélienne (Badano and Katsanis, 2002; Botstein and Risch, 2003). Leur expression dépend à la fois du contexte génétique et de facteurs environnementaux (Alcais et al., 2009). De plus, les variations génétiques touchant les régions génomiques impliquées dans la réalisation de ces phénotypes ont une pénétrance incomplète.

Deux méthodes ont été utilisées pour caractériser la composante génétique de ces phénotypes non mendéliens : les études de liaison gène candidat et les études d'associations pangénomiques (GWAS, de l'anglais *Genome-Wide Association Study*). Les approches gène candidat ont permis de mettre en évidence l'importance de nombreux gènes de l'immunité dans la réponse à certaines maladies infectieuses. C'est ainsi que plusieurs SNP du complexe majeur de l'histocompatibilité (HLA) ont été identifiés comme étant impliqués dans de nombreux troubles inflammatoires comme la sclérose en plaques (Sawcer et al., 2005). Des variations génétiques touchant d'autres gènes ont aussi été corrélées à des maladies infectieuses telles que la lèpre (Mira et al., 2004) et la tuberculose (Grant et al., 2013). L'avènement des techniques de séquençage nouvelle génération ont permis d'augmenter le nombre de variations génétiques étudiées et de procéder à des études d'associations pangénomiques. Les GWAS ont depuis été largement utilisées dans le contexte des maladies infectieuses, que ce soit dans le cadre des infections virales (Fellay et al., 2007; Ge et al., 2009; McLaren et al., 2013; Thomas et al., 2009; Troyer et al., 2011), bactériennes (Thye et al., 2012, 2010; Zhang et al., 2009) ou parasitaires (Jallow et al., 2009 ; pour une revue exhaustive des GWAS portant sur des maladies infectieuses, se rapporter à Abel et al., 2014). Cependant, les résultats obtenus par GWAS ont parfois montré les limites de ce type d'approche. Ce fut par exemple le cas pour l'étude de la susceptibilité à l'infection par *Mycobacterium tuberculosis* dont nous discuterons dans la section 4.3.

# 3. Approche génétique de l'adaptation de l'Homme aux pathogènes

Au cours de son histoire, l'Homme a dû faire face à des conditions environnementales variées. Il a notamment été confronté à une diversité importante de micro-organismes, qu'ils soient pathogéniques ou non. Le nombre croissant de données pangénomiques, telles que celles des projets HapMap et 1000 Génomes, a récemment permis d'élargir les études de génétique évolutive humaine et de génétique des populations, menant à une vision plus précise de la manière dont les pathogènes ont participé au modelage du génome humain. Dans cette partie, je discuterai de la mise en évidence des pathogènes comme principale force de sélection exercée sur le génome humain, des gènes impliqués dans la réponse immunitaire innée ainsi que de leur évolution.

## 3.1. Les pathogènes comme principale contrainte sélective

Une approche complémentaire aux études de génétique clinique et d'épidémiologie génétique, que nous avons évoquées dans la section précédente, réside dans la génétique évolutive et la génétique des populations. Cet angle d'approche permet de mettre en évidence des gènes ayant des fonctions capitales pour l'organisme et d'autres ayant participé à l'adaptation des populations humaines. Ainsi, en se rapportant à l'immunologie, ces analyses aident à la compréhension de la manière dont la sélection naturelle a mené à la diversité génétique sous-jacente aux différences de susceptibilité aux maladies infectieuses observées dans la population humaine. L'un des avantages indéniables de ces stratégies est qu'elles ne reposent pas sur des évaluations de symptômes cliniques qui sont parfois difficiles à observer et/ou évaluer. Les études de génétique évolutive et de génétique des populations permettent donc d'apporter des informations sur le fonctionnement du système immunitaire qui ne sont pas biaisées par l'étude d'un phénotype particulier.

### 3.1.1. Cas général : les études pangénomiques

Dès le début des années 2000, des études à large échelle ont cherché à mettre en évidence des signes d'une évolution particulière des gènes humains. En 2005, une comparaison des séquences nucléotidiques de 11 000 gènes entre l'homme et le chimpanzé a permis de confirmer que la sélection naturelle avait contribué aux évolutions moléculaires qui ont eu lieu dans ces deux espèces depuis leur divergence (Bustamante et al., 2005). Les auteurs ont, au cours de ces travaux, identifié plusieurs sites présentant des signes d'une évolution accélérée. Certains des gènes correspondants sont impliqués dans l'immunité impliquant les cellules NK (de l'anglais *Natural Killer*), qui présentent une activité cytotoxique envers les cellules infectées ou les cellules tumorales (voir Caligiuri (2008) pour une revue des fonctions des cellules NK). D'autres gènes codant pour des protéines aux fonctions liées à l'immunité comme les récepteurs aux immunoglobulines présentent aussi des signes de sélection positive. Une autre étude se basant cette fois sur les données HapMap (The International HapMap Consortium et al., 2003) et se focalisant sur l'adaptation des populations humaines a montré que les facteurs impliqués dans la réponse immunitaire mettant en jeu le Complexe Majeur d'Histocompatibilité (CMH) de type I étaient enrichis en signaux de sélection positive récente dans les populations européennes (Voight et al., 2006). D'autres analyses ont par la suite identifié de nombreux gènes impliqués dans les réponses

immunitaires comme faisant partie des régions géniques les plus ciblées par la sélection positive (1000 Genomes Project Consortium, 2010; Barreiro et al., 2008; Grossman et al., 2013; Pickrell et al., 2009; Sabeti et al., 2007; Wang et al., 2006). Si l'ensemble de ces études a permis de mieux apprécier certaines données épidémiologiques, comme ce fut le cas pour les différences de susceptibilité à *Plasmodium falciparum* (voir section 3.1.2), cela a aussi apporté des hypothèses inédites en proposant de nouveaux gènes candidats pouvant être impliqués dans des différences de susceptibilité aux maladies infectieuses alors que leur rôle dans l'immunité est encore peu ou pas décrit. Ainsi, *SLC24A5* se situe dans la région présentant la plus forte association avec une résistance à la lèpre (Grossman et al., 2013; Wong et al., 2010). Ce gène, principalement connu pour son implication dans la pigmentation de la peau, contient un variant génétique sous sélection positive dans les populations Européennes (dans lesquelles il est fixé) et Asiatiques (Lamason et al., 2005; Norton et al., 2007). Ces données suggèrent donc que *SLC24A5* pourrait avoir des fonctions encore inconnues dans la réponse immunitaire mise en place suite à l'infection par *Mycobacterium leprae*, fonctions qui pourraient être liées aux signaux de sélection positive détectés dans ce gène en Eurasie.

Plusieurs études ont donc identifié un grand nombre de gènes impliqués à la fois dans l'immunité innée et adaptative comme ayant participé à l'évolution de l'espèce humaine et aux adaptations de différentes populations à leur environnement.

### 3.1.2. Validation formelle des pathogènes comme principale contrainte sélective

Les données que nous venons d'évoquer suggèrent que les pathogènes ont constitué l'une des principales forces de sélection naturelle ayant guidé l'évolution de l'espèce humaine. Une étude pionnière fut menée en 2005 dans le but de tester cette hypothèse (Prugnolle et al., 2005). Les auteurs se sont intéressés à la diversité génétique retrouvée dans les régions génomiques contenant les gènes codant pour le CMH-I. En étudiant 61 populations humaines différentes, ils ont montré que les populations évoluant dans un environnement à forte diversité pathogénique présentent une diversité génétique plus importante que ce qui est attendu sous neutralité. Peu de temps après, une autre équipe a obtenu des résultats similaires en montrant qu'il existe une corrélation entre la richesse pathogénique de l'environnement et la diversité génétique touchant des gènes codant pour des antigènes du groupe sanguins et des interleukines (Fumagalli et al., 2009a, 2009b). De plus, en distinguant les différents types de

pathogènes présents dans l'environnement, les auteurs montrent que les vers parasites ont exercé une forte pression de sélection sur un ensemble de 5 gènes codant pour des interleukines, proposant que ces agents infectieux ont constitué une menace constante et stable pour les populations qui leurs sont confrontées (Fumagalli et al., 2009b).

Afin de valider définitivement l'hypothèse, une autre étude a été réalisée. La distribution spatiale des fréquences alléliques de plus de 500 000 SNP dans 55 populations ont été corrélées à des facteurs environnementaux caractéristiques du milieu de vie des populations, tels que le climat, le régime alimentaire ou la charge pathogénique (Fumagalli et al., 2011). Cette étude a révélé que la diversité locale des pathogènes est le principal facteur guidant l'adaptation des populations.

Nombre de régions ciblées par la sélection positive présentent des variations génétiques impliquées dans des maladies inflammatoires (Barreiro and Quintana-Murci, 2010; Fumagalli et al., 2011; Sabeti et al., 2007; Voight et al., 2006), ce qui suggère que ces allèles à risque ont conféré un avantage évolutif au cours de l'évolution de notre espèce. Ceci a mené à la formulation de l' « hypothèse hygiéniste », qui stipule que des allèles favorisant la réponse efficace de l'hôte face à l'infection dans un environnement où la charge pathogénique est élevée sont aujourd'hui délétères pour les organismes qui sont moins exposés aux agents infectieux. Une étude récente basée sur des données de GWAS, de scan de sélection positive, de cartographie d'eQTL et d'interaction entre protéines a récemment démontré que les SNP associés aux maladies inflammatoires sont en effet enrichis en signaux de sélection positive et touchent des gènes qui codent pour des protéines interagissant entre elles (Raj et al., 2013). De plus, le génotype à un certain nombre de ces SNP est corrélé à l'expression de transcrits, suggérant qu'ils ont participé à l'adaptation de l'Homme à son environnement sans modification de la séquence protéique (Raj et al., 2013). Cette étude démontre ainsi qu'un certain nombre de variations génétiques favorisées au cours de l'évolution de l'Homme sont aujourd'hui associées à des susceptibilités de maladies auto-immunes *via* le rôle qu'elles jouent dans le contrôle génétique de l'expression génique dans les cellules immunitaires.

### 3.1.3. L'exemple documenté de la malaria

La malaria est une maladie infectieuse touchant l'Homme depuis plus de 10 000 ans (Carter and Mendis, 2002). Elle reste aujourd'hui la maladie parasitaire la plus répandue à travers le

monde qui a causé la mort de plus de 627 000 personnes en 2012 (World Health Organization, 2013). Très tôt, des facteurs génétiques liés à la résistance à cette maladie infectieuse ont été mis en évidence. Suite aux travaux de Haldane que nous avons précédemment évoqués (voir section 2.3, Haldane, 1949), de nombreux allèles conférant une protection face à l'infection par les différentes souches de *Plasmodium* ont été identifiés. On peut par exemple citer les mutations touchant les gènes codant pour les chaines α et β de l'immunoglobuline, qui ont été associées à une plus faible susceptibilité à l'infection par *Plasmodium falciparum* (se référer à Taylor et al. (2012) pour une revue des travaux). Des études épidémiologiques et moléculaires ont aussi montré que le gène *ACKR1* (plus connu sous le nom de *DARC* pour *Duffy Antigen Receptor for Chemokines*) présente plusieurs allèles ayant tous pour conséquence de réduire fortement voire d'empêcher l'infection des érythrocytes par *Plasmodium vivax* (une revue des travaux fut effectuée par Zimmerman et al., 2013). Différentes mutations touchant le gène codant pour la G6PD et affectant l'activité de cet enzyme ont aussi été associées à des différences de susceptibilité à l'infection par ce parasite (Leslie et al., 2010; Louicharoen et al., 2009; Ruwende et al., 1995; Sirugo et al., 2014). Il semble donc que les parasites provoquant la malaria ont exercé une telle pression de sélection sur le génome humain qu'ils ont mené à une convergence évolutive illustrée par l'accumulation de plusieurs mutations distinctes dans un même gène ainsi que dans différentes séquences codantes particulièrement importantes dans le fonctionnement des cellules infectées par *Plasmodium*.

Il est important de souligner qu'un grand nombre de ces mutations ont été associées à des maladies mendéliennes. C'est par exemple le cas de l'allèle S de la chaîne $\beta$ de l'hémoglobine qui confère une protection face à l'infection à l'état hétérozygote mais qui, à l'état homozygote, provoque la drépanocytose et peut être létale en l'absence de traitement médical adéquat. Des déficiences en G6PD ont, elles, été prouvées comme étant les causes de certaines anémies hémolytiques aiguës déclenchées par une infection (Choremis et al., 1966) et/ou la prise de certains médicaments (Cappellini and Fiorelli, 2008). Ainsi, la protection envers la malaria a un certain coût pour l'organisme, laissant penser à un équilibre précaire entre protection envers la maladie infectieuse et danger pour la survie de l'individu. Cependant, nombre de régions contenant des allèles permettant à l'organisme de se prémunir contre l'infection par *Plasmodium* montrent des signes de sélection positive. C'est notamment le cas pour *CD36* (Fry et al., 2009), *G6PD* (Louicharoen et al., 2009; Sabeti et al., 2002; Tishkoff et al., 2001) ou encore *ACKR1* (Hamblin and Di Rienzo, 2000).

L'ensemble de ces données illustre là encore le fait que certains pathogènes ont été les principaux facteurs influençant l'évolution du génome des populations humaines.

## 3.2. Mécanismes de défense de l'hôte humain contre les pathogènes

### 3.2.1. Immunités innée et adaptative

Les pathogènes constituant une menace permanente pour tous les organismes, les êtres vivants ont développé de nombreux mécanismes de défense qui sont activés par les agents infectieux et qui participent à la protection de l'hôte en détruisant les micro-organismes invasifs et en neutralisant leurs facteurs de virulence. En se basant sur le type de récepteurs utilisés pour reconnaitre les micro-organismes, deux branches de de la réponse immunitaire ont été définies : l'immunité innée et l'immunité adaptative. Les récepteurs utilisés par la première sont codés dans la lignée germinale et permettent de définir un système de défense évolutivement ancien (Hoffmann et al., 1999; Janeway and Medzhitov, 2002; Litman et al., 2005). L'immunité adaptative, aussi dite acquise, n'est retrouvée que chez les vertébrés. Elle repose sur l'utilisation d'un panel très large de récepteurs codés dans la lignée somatique (Janeway, 2001; Medzhitov and Janeway, 1998).

Pendant longtemps, les travaux portant sur le système immunitaire ont été largement guidés par la compréhension des mécanismes de l'immunité adaptative. Ils ont permis de mettre en évidence le rôle des lymphocytes B et T et de leurs récepteurs aux antigènes spécifiques (Janeway, 2001). Au cours de leur développement, des évènements de recombinaison et de variation de segments de gènes génèrent un récepteur à un antigène unique par lymphocyte qui possède un site de reconnaissance particulier. Si cela permet d'avoir un réservoir quasiment illimité de récepteurs et d'anticorps, les lymphocytes ne deviennent efficaces que deux à quatre jours après le premier contact avec le micro-organisme. Cette période correspond au temps nécessaire à l'expansion clonale des lymphocytes dont les récepteurs reconnaissent l'un des antigènes de l'agent infectieux en question. Malgré tout, la réponse induite est hautement spécifique et les lymphocytes persistent ensuite sous forme d'une mémoire immunitaire qui assure une protection plus rapide en cas de nouvelle exposition au pathogène considéré. Etant donné que le temps de génération de nombreux agents infectieux est de l'ordre de la dizaine de minutes, il apparait évident qu'un mécanisme de réponse immédiate est nécessaire.

Plus récemment, des travaux ont commencé à mettre en évidence le rôle du système immunitaire inné dans l'activation de la réaction inflammatoire et de la réponse immunitaire adaptative. Cette branche de l'immunité commence au niveau des zones d'interface entre le milieu intérieur des organismes pluricellulaires et le milieu extérieur en mettant en jeu des peptides antimicrobiens qui permettent de se prémunir contre les infections bactériennes et fongiques (Braff et al., 2005; Ganz and Lehrer, 1994; Wang, 2014; Zasloff, 2002). Dans le cas d'un franchissement des barrières épithéliales par un micro-organisme, les phagocytes, les lymphocytes NK et le système du complément entrent en jeu. Les protéines du complément circulent dans le sang où elles peuvent être activées selon différentes voies pour détruire les pathogènes, faciliter la phagocytose, recruter des cellules inflammatoires ou modeler la réponse immunitaire adaptative (Ricklin et al., 2010). Les phagocytes et les cellules NK sont activés par la reconnaissance de motifs moléculaires conservés au sein des micro-organismes et absents des cellules de l'hôte appelés *Pathogen Associated Molecular Pattern* (PAMP) (Janeway and Medzhitov, 2002; Medzhitov, 2001) ou *microbial sensors* (Beutler et al., 2006). Les récepteurs des différents PAMP sont appelés *Pathogen Recognition Receptors* (PRR). Leur fixation à un PAMP va induire l'activation de voies de signalisations particulières, aboutissant à une réponse immunitaire (Janeway, 1989). Il existe une grande variété de PRR. Ils peuvent être localisés à la membrane des cellules, dans le cytoplasme, ou être sécrétés. Plusieurs familles de PRR ont été définies en se basant sur la structure des protéines.

### 3.2.2. Les différentes familles de PRR

#### 3.2.2.1. Les récepteurs de type TOLL

Les récepteurs de types TOLL (TLR pour *Toll-like receptors*) ont été les premiers PRR identifiés et ont été largement étudiés depuis. Ils présentent de fortes homologies de structure avec la protéine codée par le gène *Toll*, identifiée pour la première fois chez *Drosophila melanogaster* comme étant impliquée dans le développement (Anderson et al., 1985; Hashimoto et al., 1988). Ce n'est qu'à partir de 1996 que cette protéine fut démontrée comme jouant un rôle clé dans la protection de l'hôte face aux infections par des bactéries à Gram positif ou des champignons (Hoffmann, 2003; Lemaitre et al., 1996; Lemaitre and Hoffmann, 2007). Par la suite, une famille de dix TLR a été identifiée chez l'Homme (Chuang and Ulevitch, 2000, 2001; Du et al., 2000; Medzhitov et al., 1997; Rock et al., 1998; Takeuchi et

al., 1999). Quatre d'entre eux, TLR3, TLR7, TLR8 et TLR9, sont localisés dans des compartiments intracellulaires comme les endosomes alors que TLR1, TLR2, TLR4, TLR5, TLR6 et TLR10 sont exprimés à la surface cellulaire (Akira et al., 2006; Takeda et al., 2003). Les TLR intracellulaires sont principalement impliqués dans la reconnaissance d'acides nucléiques alors que les récepteurs de la membrane plasmique détectent des produits dérivés d'une large variété de bactéries, de parasites et de champignons (Akira et al., 2006; Blasius and Beutler, 2010; Kawai and Akira, 2011, 2006; Takeda et al., 2003). Il faut noter que le ligand de TLR10 n'est pas encore déterminé et que ses fonctions ne sont que peu documentées (Godfroy et al., 2012; Guan et al., 2010; Hasan et al., 2005; Hornung et al., 2002; Oosting et al., 2014).

La fixation d'un PAMP sur un TLR induit l'activation d'une voie de signalisation. La première étape correspond au recrutement d'adaptateurs contenant un domaine TIR. MYD88 peut être recruté par tous les TLR à l'exception de TLR3, TIRAP (parfois appelé MAL) est recruté par TLR2 et TLR4, TICAM1 (TRIF) par TLR3 et TLR4 et TIVAM2 (TRAM) n'est utilisé que par TLR4 (Akira and Takeda, 2004; Kawai and Akira, 2010; Kumar et al., 2011; Takeda et al., 2003). La transmission du signal par de nombreux intermédiaires moléculaires aboutit à l'activation des facteurs de transcription NF-κB et AP1 qui permettent la production de cytokines inflammatoires telles que l'IL-1β, l'IL-6 ou le TNFα (Kawai and Akira, 2011; Lee and Kim, 2007). L'activation des TLR endosomaux entraîne l'activation de facteurs comme IRF3 et IRF7, menant à la production d'interférons de type I. Cette voie de signalisation peut aussi être utilisée par TLR4, dans le cadre d'une signalisation indépendante de MYD88 (Akira and Takeda, 2004; Blasius and Beutler, 2010; Kawai and Akira, 2011).

#### 3.2.2.2. Les lectines de type C

Les lectines de type C, aussi appelées CLR pour l'anglais *C-type lectins,* constituent une grande classe de protéines solubles ou transmembranaires. Elles tirent leur nom du fait que, chez les vertébrés, ces récepteurs fixent de nombreux glucides (activité lectine) par l'intermédiaire d'un domaine particulier, le CRD (*carbohydrate recognition domain*). Cette interaction nécessite la présence d'ions $Ca^{2+}$. L'ensemble des CLR présente au moins un domaine CRD ou un domaine homologue qui ne fixe pas nécessairement les glucides. Ce domaine protéique détermine les structures spécifiques qui sont reconnues par le récepteur (Veldhuizen et al., 2011; Zelensky and Gready, 2005). En se basant sur des similarités de

structures et de fonctions, 17 familles de CLR ont été définies. Si les CLR participent à la reconnaissance d'agents infectieux (Drummond and Brown, 2013; Graham and Brown, 2009; Huysamen and Brown, 2009; Kanazawa, 2007; Kanazawa et al., 2004), elles permettent aussi une réponse à des signaux de dangers cellulaires comme la présence de lipides oxydés, participant ainsi au maintien de l'intégrité tissulaire.

Les CLR transmembranaires peuvent être divisées en deux grands groupes : les CLR de type Dectin-1 et les CLR de type Dectin-2. Les récepteurs de type Dectin-1 possèdent un CRD dans leur partie extracellulaire, un domaine transmembranaire et un domaine de transduction du signal dans leur partie intracellulaire (Huysamen and Brown, 2009). Certains membres de ce groupe présentent un motif ITAM, résultant en une activation de facteurs de transcription lors de la fixation des ligands sur ces récepteurs. D'autres de ces senseurs présentent un motif ITIM dans leur partie intracellulaire, qui participe à l'inhibition des facteurs de transcription en aval de la voie de signalisation. Les CLR de type Dectin-2, aussi appelées CLR classiques, ont, elles, une partie intracellulaire particulièrement restreinte qui ne contient pas de domaine d'activation ou d'inhibition. Afin de transmettre le signal à la cascade moléculaire, elles nécessitent donc un partenaire protéique. La fixation du ligand sur ces récepteurs est donc suivie du recrutement de la chaine gamma du récepteur Fc qui transmet le signal *via* son domaine ITAM (Cao et al., 2007; Kanazawa et al., 2003).

L'activation des voies CLR peut avoir des conséquences variées. D'une part, la reconnaissance de certains PAMP par ces PRR induit une signalisation via la voie impliquant CARD9, BCL10 et MALT1, menant à l'activation de NF-κB et à la production de cytokines inflammatoires telles que l'IL-6, l'IL-1β ou le TNFα (Kerrigan and Brown, 2011). De récents travaux ont aussi montré que l'activation de certains CLR pouvait induire une activation de l'inflammasome impliquant la caspase-8 (Gringhuis et al., 2012). Finalement, les CLR permettent d'amorcer l'internalisation des pathogènes ou des corps étrangers par les phagocytes.

#### 3.2.2.3. Les récepteurs de type RIG-I

Les récepteurs de type RIG-I, aussi appelés RLR pour *Rig-I-like Receptors*, ont commencé à être identifié au début des années 2000. Le premier récepteur mis en évidence, RIG-I (aujourd'hui appelé DDX58) (Yoneyama et al., 2004), a donné son nom à la famille qui contient deux autres protéines : IFIH1 (parfois appelé MDA5), et DHX58 (LGP2). Ces trois molécules reconnaissent des ARN viraux dans le cytoplasme et participent à la réponse

antivirale (Baum and García-Sastre, 2010; Loo and Gale, 2011; Yoneyama et al., 2005, 2004). Ces récepteurs ont en commun un domaine servant à la reconnaissance des ARN et conférant leur spécificité aux récepteurs, un domaine hélicase permettant l'intégration de signaux induits par les ARN (Luo et al., 2013) et deux domaines CARD (*Caspase Recruitment Domains*) déclenchant la réponse inflammatoire et, dans certains cas, la mort cellulaire (Kang et al., 2002; Kovacsovics et al., 2002; Saito et al., 2007; Yoneyama et al., 2005). La protéine DHX58 ne possède pas ces derniers domaines. Elle pourrait donc réguler l'activité de DDX58 et IFIH1 par (i) inhibition compétitive, en fixant les ARN double-brins et en réduisant leur disponibilité, (ii) interaction entre le domaine répresseur de DHX58 et le domaine CARD de DDX58 ou (iii) inhibition compétitive par recrutement de l'adaptateur nécessaire à la transduction du signal suite à l'activation de DDX58. Cependant, il semblerait que DHX58 participe à l'activation de IFIH1 (Venkataraman et al., 2007).

L'activation de DDX58 et IFIH1 induit le recrutement de MAVS qui va, par l'intermédiaire d'autres partenaires moléculaires, mener à l'activation d'IRF3 et IRF7 pour induire l'expression des gènes codant pour les interférons de type I (Dixit et al., 2010; Hiscott et al., 2006; Lin et al., 2006; Ohman et al., 2009) ou à l'activation de NF-κB qui participera à l'expression de cytokines pro-inflammatoires (Kato et al., 2005; Kawai et al., 2005; Poeck et al., 2010). D'autres données indiquent que l'activation de la voie des RLR peut induire la production d'interféron de type III (IFNλ) (Donnelly and Kotenko, 2010; Loo and Gale, 2011; Onoguchi et al., 2007; Osterlund et al., 2005).

#### 3.2.2.4. Les récepteurs de type NOD

Les récepteurs de type NOD, ou NLR (*NOD-like Receptors*) sont retrouvés chez différents vertébrés et invertébrés (Lange et al., 2011; Rast et al., 2006; Tian et al., 2009). Des senseurs ayant une structure similaire sont aussi retrouvés chez les plantes. Ces NBS-LRR végétaux (*Nucleotide-Binding Site Leucine-Rich Repeats*) sont des cibles de la sélection positive (Meyers et al., 1998; Mondragón-Palomino et al., 2002; Parniske et al., 1997) et participent aux mécanismes clés de l'immunité des plantes (Caplan et al., 2008). Chez l'Homme, les NLR sont au nombre de 22. Principalement localisés dans le cytoplasme même si certains de ces senseurs sont retrouvés dans des compartiments cellulaires (Moore et al., 2008; Steimle et al., 1993), ils participent à la reconnaissance de structures moléculaires microbiennes et de signaux endogènes de danger ou de stress (Kufer and Sansonetti, 2011). Ils présentent une

structure protéique similaire avec (i) un domaine LRR qui sert à la fixation du ligand, (ii) un domaine NACHT qui permet l'oligomérisation de la protéine et (iii) un domaine effecteur en partie N-terminale. Les variations que présente ce dernier domaine permettent de distinguer la sous-famille des NOD (*Nucleotide Oligomerization Domain*) de celle des NALP (*NACHT, Leucine-rich repeat, Pyrin domain*).

Les NOD reconnaissent des composants bactériens et activent le facteur de transcription NF-κB *via* les adaptateurs RIPK2 et CARD9, ce qui a pour effet d'induire l'expression de cytokines pro-inflammatoires (Girardin et al., 2001; Hsu et al., 2007; Inohara et al., 2000; Kobayashi et al., 2002; Ogura et al., 2001). Les NALP, eux, font partie d'un complexe protéique appelé inflammasome. Par l'intermédiaire de l'adaptateur PYCARD (parfois appelé ASC ou CARD5), l'activation des NALP va induire la stimulation de l'activité protéolytique de CASP1 qui va cliver les pro-IL-1β et pro-IL18, permettant la production de la forme mature de ces cytokines (Mariathasan et al., 2006; Martinon et al., 2002; Schroder and Tschopp, 2010).

#### 3.2.2.5. Autres senseurs des acides nucléiques : les CNAS

Si nous avons vu que les RLR participaient à la reconnaissance d'acides nucléiques localisés dans le cytosol, il est parfois possible que DDX58 ne soit pas le premier acteur moléculaire de la voie de signalisation. En effet, il a récemment été montré que ce senseur reconnait des ARN intermédiaires (Ablasser et al., 2009; Chiu et al., 2009) qui ont été transcrits par l'ARN polymérase III à partir d'ADN double-brins riches en AT (Figure 8). Cette polymérase peut donc jouer le rôle de senseur dans les cellules humaines.

Hormis les RLR et les NLR, d'autres récepteurs appelés CNAS (pour *Cytosolic Nucleic Acid Sensors*) participent à la reconnaissance d'acides nucléiques présents dans le cytosol (revus dans Dempsey and Bowie, 2015; Vabret and Blander, 2013). La mise en évidence de ces mécanismes est récente puisqu'elle a commencé en 2008 avec l'identification d'une nouvelle protéine particulièrement importante : STING. Cette molécule joue un rôle dans la réponse aux agents infectieux viraux et bactériens et dans la reconnaissance des ADN de l'hôte au cours de maladies auto-immunes (Burdette and Vance, 2013; Ishikawa and Barber, 2008; Zhong et al., 2008). Une fois activé par son ligand, STING recrute TBK1 qui va phosphoryler IRF3, induisant une réponse impliquant les interférons de type I (Ishikawa et al., 2009; Stetson and Medzhitov, 2006; Tanaka and Chen, 2012).

STING étant capable de reconnaitre certains métabolites secondaires des bactéries comme le diguanylate monophosphate cyclique (c-di-GMP) (Burdette et al., 2011), il a récemment été proposé qu'il participe aussi à la transduction du signal initié par d'autres protéines. Ainsi, cGAS (*cyclic-GMP-AMP synthase*) reconnait certains ADN bactériens et viraux (Li et al., 2013; Schoggins et al., 2014) et génère un messager secondaire, cGAMP (Sun et al., 2013), qui va à son tour participer à l'activation de STING et à la production d'interférons de type I (Zhong et al., 2008).

STING est aussi utilisé comme adaptateur par IFI16 une fois que celui-ci se lie à des ADN ou ARN microbiens (Horan et al., 2013; Unterholzner et al., 2010). Cette dernière protéine présente des homologies de structure avec un autre PRR reconnaissant des ADN double-brins viraux dans le cytoplasme : AIM2. Ce récepteur, une fois lié à son ligand, induit l'activation de la caspase 1 *via* le recrutement de PYCARD (Hornung et al., 2009). IFI16 et AIM2 possèdent toutes deux un domaine PYRIN ainsi qu'un site de liaison à l'ADN (HIN200), ce qui a permis de définir une nouvelle classe de PRR appelés ALR (*AIM2-Like*

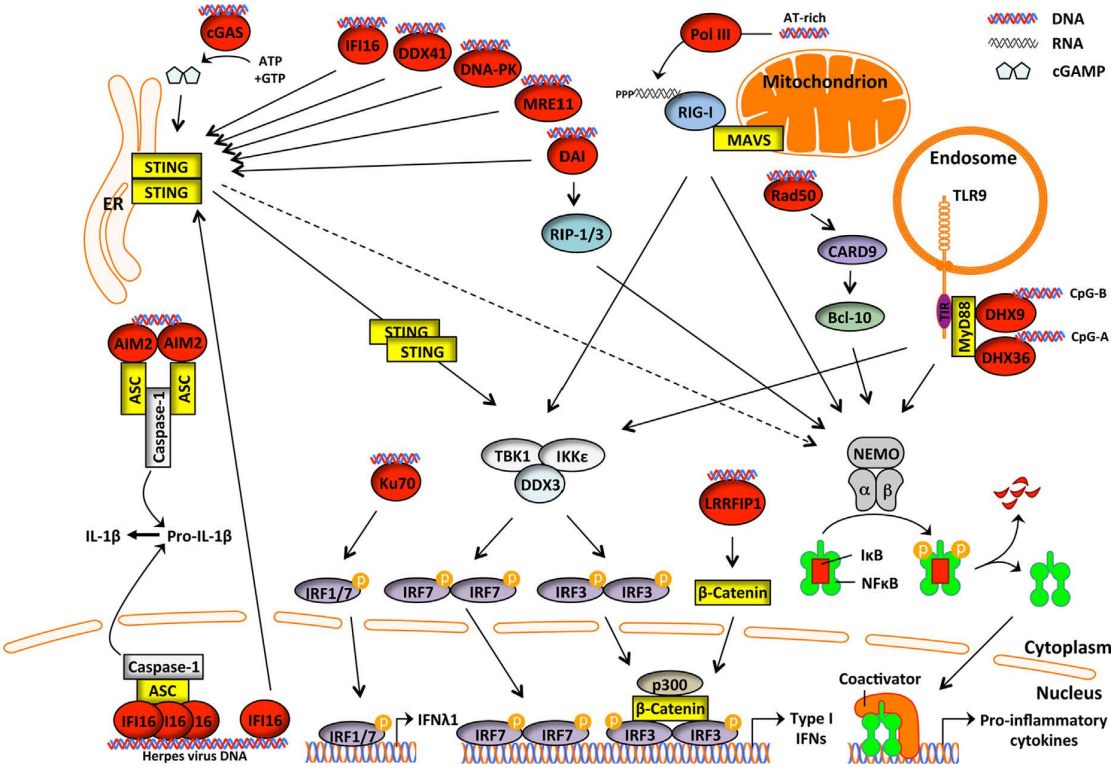



**Figure 8. Senseurs d'ADN cytosoliques et voies de signalisation.**
Tiré de Dempsey and Bowie (2015). La reconnaissance des ADN par les senseurs spécifiques engendre une cascade de signalisation qui aboutit à la production de cytokines pro-inflammatoires, d'IFN et/ou à l'activation de l'IL1β.

*Receptors*) qui contient aussi MNDA et PYHIN1 (Keating et al., 2011; Schattgen and Fitzgerald, 2011).

La mise en évidence récente de nouveaux PRR montre que nous sous-estimons très certainement le nombre de molécules impliquées dans la reconnaissance de PAMP et laisse supposer que de nombreux mécanismes de défense face aux pathogènes restent encore inconnus à ce jour.

## 3.3. Génétique évolutive de l'immunité innée

Comme nous l'avons précédemment évoqué, les pathogènes ont été l'une des principales menaces quant à la survie et la procréation de l'Homme. Etant donné que le bagage génétique d'un individu influence fortement sa susceptibilité aux maladies infectieuses, les pressions sélectives exercées par les agents infectieux ont fortement participé au modelage du génome humain. L'immunité innée constituant la première ligne de défense de l'hôte contre les pathogènes, les gènes codant pour les protéines impliquées dans ce processus cellulaire constituent un modèle d'étude particulièrement relevant pour étudier les pressions sélectives exercées par les micro-organismes sur le génome humain.

Les premières analyses des traces moléculaires laissées sur l'ensemble du génome par la sélection naturelle ont montré que certains gènes impliqués dans l'immunité ont été la cible de mutations récurrentes ayant conféré un avantage sélectif aux primates (Bustamante et al., 2005; Kosiol et al., 2008; Nielsen et al., 2005a). Ce processus cellulaire fait aussi partie des fonctions cellulaires présentant le plus de signatures de sélection positive très ancienne (Clark et al., 2003). De même, de nombreuses études intraspécifiques ont mis en évidence des régions génomiques associées à l'immunité comme étant positivement sélectionnées dans des

populations humaines distinctes (Barreiro et al., 2008; Pickrell et al., 2009; Sabeti et al., 2007; Voight et al., 2006). Cependant, une comparaison entre les séquences de plus de 11 000 gènes entre l'Homme et le Chimpanzé réalisée au début du siècle a révélé qu'un grand nombre de gènes impliqués dans la réponse à l'infection ont un rôle fonctionnel particulièrement important pour l'organisme puisqu'ils n'accumulent que très peu de mutations modifiant les séquences d'acides aminés pour lesquelles ils codent (Bustamante et al., 2005). Il est donc apparu dès le début des études de génétique évolutive considérant les gènes de l'immunité que ceux-ci étaient soumis à des forces sélectives variées. De

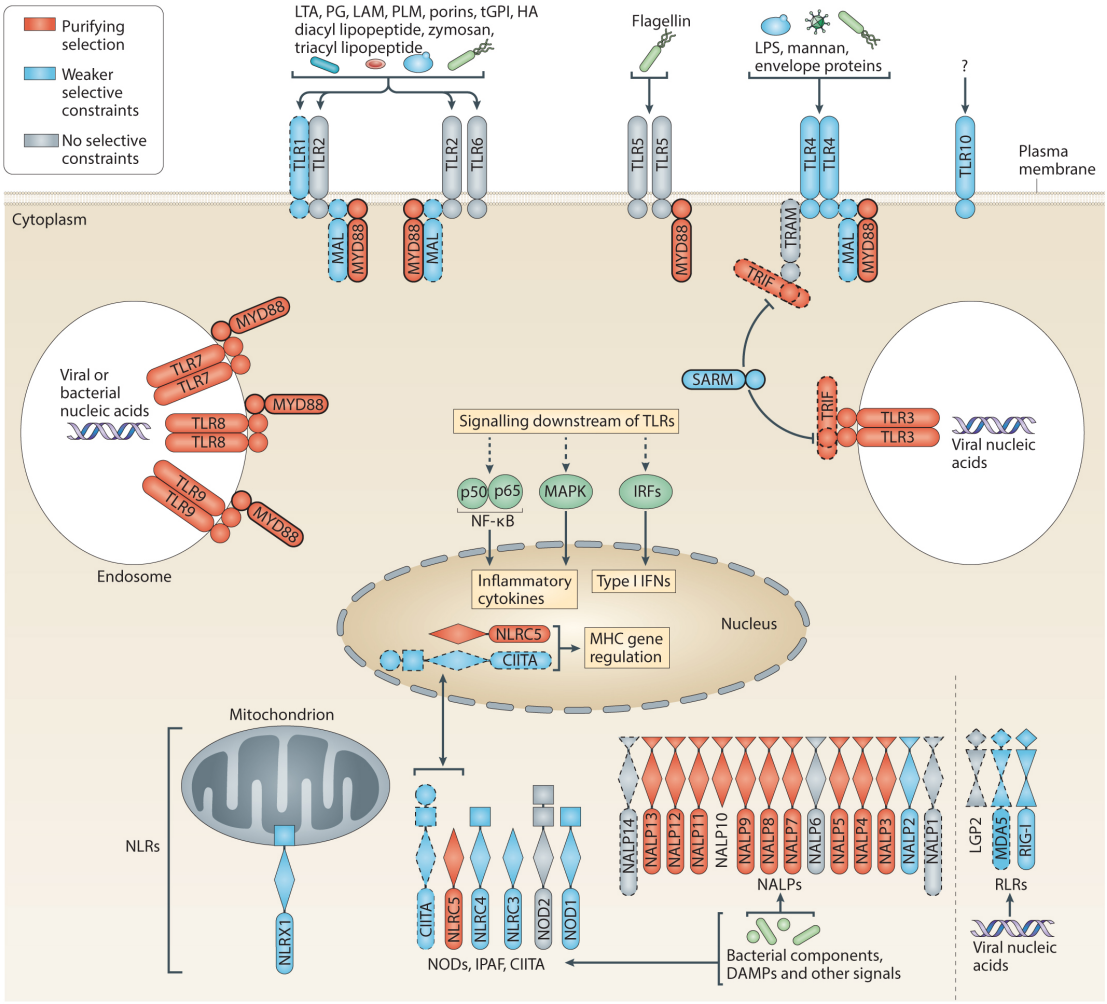



**Figure 9. Pressions de sélection exercées sur certains récepteurs et adaptateurs de l'immunité innée.**
Tiré de Quintana-Murci and Clark (2013). Les gènes ayant évolué sous l'action de la sélection purificatrice sont montrés en rouge, ceux ayant évolué sous sélection négative plus relâchée sont indiqués en bleu, ceux pour lesquels aucun signe de contrainte sélective n'a été trouvé sont montrés en gris. En gras sont indiqués les gènes pour lesquels des évidences de sélection positive au niveau de l'espèce humaine ont été retrouvées et en pointillés ceux présentant des signes d'une sélection positive restreinte à certaines populations humaines.

nombreuses études portant sur des gènes ou des familles de gènes candidat(e)s ont alors été menées dans le but d'identifier les forces sélectives ayant modelé les gènes impliqués dans la réponse à l'infection ainsi que d'en estimer les intensités (Figure 9, Quintana-Murci and Clark, 2013).

### 3.3.1. Évolution des TLR

Les TLR sont, globalement, soumis à l'action de la sélection purificatrice (Barreiro et al., 2009; Smirnova et al., 2001; Wlasiuk and Nachman, 2010). Malgré tout, l'intensité des pressions de sélection exercées sur ces dix senseurs n'est pas homogène.

Ainsi, les TLR localisés dans les endosomes n'accumulent que très peu de mutations non-synonymes et sont soumis à une contrainte évolutive plus importante que les TLR localisés à la membrane cellulaire (Barreiro et al., 2009). L'une des hypothèses avancées pour expliquer cette structure porte sur les motifs reconnus par les TLR localisés dans les compartiments cellulaires. Ceux-ci ont pour principaux ligands les acides nucléiques de micro-organismes qui, bien que caractérisés par certaines modifications chimiques (Karikó et al., 2005), sont suffisamment ressemblants aux acides nucléiques de l'hôte pour qu'une modification du site de fixation du récepteur induise une reconnaissance du matériel génétique de l'hôte et mène à une réponse auto-immune (Krieg and Vollmer, 2007; Marshak-Rothstein, 2006; Pisitkun et al., 2006). Ainsi, la conservation des TLR endosomaux pourrait refléter la balance entre reconnaissance efficace des ARN du « non-soi » et non fixation du matériel génétique de l'hôte.

Les TLR localisés à la surface cellulaire présentent plus de flexibilité quant à l'acceptation de mutations modifiant les séquences d'acides aminés, ainsi qu'une plus grande diversité génétique et fonctionnelle entre populations. Ainsi, jusqu'à 23 % des individus portent une variation génétique induisant une modification potentiellement dommageable pour la protéine dans l'un des gènes codant ces senseurs et jusqu'à 16 % des membres d'une population peuvent porter une mutation non-sens dans l'une de ces séquences (Barreiro et al., 2009; Wlasiuk et al., 2009). Ces données suggèrent qu'il existe un certain degré de redondance au niveau des TLR participant à la reconnaissance des bactéries, champignons et protozoaires au niveau de la membrane cellulaire.

Des signatures moléculaires de l'action de la sélection positive ont aussi été identifiées dans certains gènes codant pour des TLR. Il a par exemple été montré que TLR1 et TLR4 sont soumis à une évolution accélérée ancienne chez les primates, ayant entrainé d'importantes divergences de séquences entre les espèces considérées (Nakajima et al., 2008; Ortiz et al., 2008; Wlasiuk and Nachman, 2010). Chez l'Homme, une mutation non-synonyme localisée dans le gène *TLR5* régulant l'activation de NF-κB est sous sélection positive récente chez les Yoruba (Grossman et al., 2013). Une mutation non-sens est aussi retrouvée dans ce gène chez les populations européennes à une fréquence de 10 % (Barreiro et al., 2009). De même, le cluster *TLR6-1-10* est la cible d'évènements de sélection positive dans les populations Européennes (Barreiro et al., 2009; Pickrell et al., 2009). Des études fonctionnelles ont montré qu'une des mutations non-synonymes portées par l'haplotype retrouvé à forte fréquence dans ces populations participe à la régulation de l'activité du facteur de transcription NF-κB suite à la stimulation de la voie TLR1 et participe à la protection contre l'infection par *Mycobacterium Leprae* (Barreiro et al., 2009; Ben-Ali et al., 2011; C. M. Johnson et al., 2007; Misch et al., 2008). Ces données suggèrent qu'une réduction de l'activité de NF-κB pourrait avoir conféré un avantage sélectif aux individus porteurs de ces mutations. Finalement, d'autres travaux indiquent que le cluster *TLR6-1-10* est aussi sujet à l'action de la sélection positive chez le Chimpanzé et l'Orang-Outan (Enard et al., 2010). Cela suggère que cette région génomique est un point chaud de sélection darwinienne portant des variations génétiques ayant conféré un avantage sélectif à de nombreux primates.

### 3.3.2. Évolution des CLR

Très peu d'études ont porté sur l'analyse des forces sélectives guidant l'évolution des gènes codant pour les lectines de type C. L'histoire sélective de CD209 et CD209L a notamment été étudiée. Alors que le gène codant pour le premier récepteur n'accumule que très peu de mutations non-synonymes, signe d'une très forte contrainte sélective, le second présente des signes de sélection balancée ou de relâche de contrainte sélective dans les populations non africaines. D'importants débats ont concerné un autre CLR particulier : le récepteur soluble au mannose MBL2. De nombreux allèles engendrant une déficience en MBL2 peuvent être retrouvés dans diverses populations réparties sur l'ensemble du globe (Bernig et al., 2004; J.-L. Casanova and Abel, 2004; Verdu et al., 2006), suggérant qu'une perte de la fonction de ce

gène pourrait constituer un avantage sélectif. Cependant, les déficiences en MBL2 ont été associées à une plus grande susceptibilité à de nombreuses maladies infectieuses (Hibberd et al., 1999; Hoal-Van Helden et al., 1999; Mullighan et al., 2002; Peterslund et al., 2001; Roy et al., 2002; Santos et al., 2001). Une hypothèse conciliant ces deux constats semble compliquée à établir. Le débat a aussi porté sur les forces sélectives ayant modelé *MBL2*. En effet, si certaines études de génétique des populations ont proposé que ce locus était ciblé par la sélection balancée (Bernig et al., 2004; Mukherjee et al., 2009), d'autres suggèrent que ce gène évolue sous neutralité (Verdu et al., 2006). Il semblerait donc que le rôle de MBL2 soit redondant et que d'autres facteurs moléculaires soient d'une importance fonctionnelle plus grande.

### 3.3.3. Évolution des RLR et des NLR

Les NLR et les RLR, comme nous l'avons vu précédemment, participent à la reconnaissance de pathogènes et de signaux de dangers dans le cytoplasme. La famille des RLR semble jouer un rôle relativement redondant par rapport aux autres familles de senseurs puisque l'ensemble de ses membres évolue sous le régime d'une faible sélection négative, voire de neutralité (Vasseur et al., 2012). Cependant, l'intensité de cette faible pression de sélection n'est pas homogène parmi les trois RLR. Moins de mutations altérant la structure protéique sont accumulées dans DDX58 comparé à IFIH1 et DHX58 (Vasseur et al., 2011). Ce motif pourrait être la conséquence de la reconnaissance d'un nombre plus important d'ARN par DDX58 que par les deux autres protéines ou de la plus grande exigence requise par la reconnaissance de ses substrats microbiens (Loo and Gale, 2011). De plus, des indices moléculaires laissent penser que des mutations non synonymes localisées dans *IFIH1* et *DHX58* sont sous sélection positive dans des populations humaines spécifiques (Fumagalli et al., 2010; Vasseur et al., 2011) et ont conféré un avantage sélectif aux individus les portant.

L'évolution des NLR est elle aussi guidée par des évènements sélectifs de natures et d'intensités variées. Ainsi, les mutations non-synonymes touchant les gènes codant pour les NALP semblent être particulièrement délétères pour l'hôte puisque ces gènes en présentent une faible proportion par rapport à ce qui est attendu sous hypothèse de neutralité (Vasseur et al., 2012). À l'inverse, le rôle fonctionnel des récepteurs NOD parait redondant puisqu'ils évoluent sous faible sélection négative. Comme pour les autres familles de gènes que nous avons mentionnées précédemment, certains membres de la famille des NLR ont participé à

l'adaptation de populations humaines à leur environnement. Des évènements de sélection positive ont notamment ciblé les gènes *NALP1*, *NALP14* et *CIITA* (Vasseur et al., 2012).

### 3.3.4. Évolution des CNAS

Comme nous l'avons précisé dans la partie 3.2.2.5, l'identification des ALR est particulièrement récente. Les études de génétique évolutive s'étant intéressées à cette classe de PRR sont donc particulièrement restreintes pour le moment. Les premières données portant sur ces senseurs indiquent qu'ils sont très diversifiés chez les mammifères (Brunette et al., 2012). D'autres travaux ont montré qu'*IFI16* présente un plus grand nombre de substitutions non-synonymes que de mutations synonymes chez les primates (Cagliani et al., 2014). Des évidences moléculaires de l'action d'une sélection positive ancienne sur *AIM2* ont aussi été retrouvées au cours de cette étude. Deux des sites de AIM2 et un d'IFI16 sélectionnés positivement sont localisés à proximité du domaine de fixation à l'ADN, suggérant qu'ils participent à la spécificité de substrat de ces récepteurs. La région séparant les deux domaines HIN200 de IFI16 présente aussi diverses évidences d'une action de la sélection balancée sur cette région chez l'Homme (Cagliani et al., 2014). L'ensemble de ces résultats attestent de l'implication des ALR dans l'adaptation des espèces et des populations à leur environnement microbien.

### 3.3.5. Évolution de molécules autres que les PRR

Les études de génétique évolutive portant sur les gènes de l'immunité innée ne se sont pas restreintes aux senseurs des micro-organismes. Il a notamment été montré que les adaptateurs utilisés pour la transduction du signal entre les TLR et le reste de la voie de signalisation, à savoir MYD88, TIRAP, TICAM1 et TICAM2 ont été soumis à une forte contrainte sélective chez l'ensemble des primates (Nakajima et al., 2008), soulignant l'importance fonctionnelle de ces molécules. Chez l'Homme, MYD88 et TRIF présentent de très faibles taux de mutations non-synonymes au regard des variations ne modifiant pas la séquence protéique, ce qui souligne le rôle non redondant qu'elles ont dans la transmission du signal de reconnaissance des agents infectieux (Fornarino et al., 2011). Des traces moléculaires d'un balayage sélectif sont aussi retrouvées dans *MYD88* pour l'ensemble de la population humaine. Quant aux trois autres adaptateurs, ils présentent des évidences de sélection positive

dans des populations spécifiques, suggérant qu'ils ont participé à l'adaptation de groupes d'individus à un environnement particulier.

Des études de génétique évolutive ont aussi mis en évidence les différentes voies évolutives empruntées par les gènes codant pour les interférons (IFN) (Manry et al., 2011a, 2011b). *IFNγ* n'accepte aucune mutation non-synonyme, corroborant des observations cliniques du rôle non redondant de cette cytokine dans la réponse immunitaire et de son importance pour la survie de l'hôte (Manry et al., 2011b; Zhang et al., 2008). Si les deux autres gènes impliqués dans la voie de signalisation initiée par l'IFNγ peuvent accumuler des mutations non-synonymes, ils évoluent malgré tout sous la pression de la sélection négative (Manry et al., 2011b). Comme pour le groupe des IFN de classe II, l'intensité de la sélection négative guidant l'évolution des gènes codant pour les IFN de classe II est variable au sein de cette sous-famille (Manry et al., 2011a). Ainsi, certains IFN de type I comme IFNα6, IFNα8, IFNα13 et IFNα14 ne présentent que peu de changements d'acides aminés entre les populations humaines. À l'inverse, d'autres IFN de type I ont accumulé des mutations non-synonymes ou des mutations stop qui sont aujourd'hui retrouvées à des fréquences alléliques élevées dans la population, suggérant que leurs fonctions sont redondantes. Finalement, des mutations présentes dans les gènes codant pour les IFN de type III ont été ciblées par la sélection positive dans la population européenne. Ces gènes semblent donc avoir participé à l'adaptation de ces individus à leur environnement pathogénique.

Finalement, une dernière classe d'effecteurs de l'immunité a été étudiée : les défensines. Des analyses interspécifiques ont révélé que les gènes codant pour les défensines de type α et β ont évolué en étant soumis à diverses pressions de sélection positives et négatives (Aldred et al., 2005; Cagliani et al., 2008; Das et al., 2010; Hollox and Armour, 2008). Le cas de *DEFB1* est particulièrement intéressant puisque les motifs moléculaires qui lui sont associés laissent supposer l'action d'une sélection balancée à long terme (Cagliani et al., 2008).

L'ensemble de ces données montre la diversité des forces évolutives et l'intensité de leur action sur les gènes impliqués dans la réponse immunitaire innée.

# 4. Combinaison d'approches génétiques et moléculaires pour étudier les différences de réponse immunitaire

Nous avons pour l'instant principalement évoqué les variations génétiques touchant les séquences codantes. Cependant, le génome humain est principalement constitué de séquences intergéniques qui sont entre autres impliquées dans la régulation de l'expression des gènes. Il a régulièrement été proposé que cette modulation pourrait avoir joué un rôle majeur dans l'adaptation des populations à leur environnement (ENCODE Project Consortium, 2012; Grossman et al., 2013; King and Wilson, 1975; Wray, 2007). Dans cette partie, je discuterai du contrôle génétique de l'expression des gènes, de la régulation de cette expression par les miARN ainsi que d'un exemple particulier, la tuberculose.

## 4.1. Contrôle génétique de l'expression des gènes

### 4.1.1. Principe des eQTL et complémentarité avec les GWAS

Une variation génétique peut faire varier un trait phénotypique si elle est située dans une séquence codante, perturbant la séquence d'acides aminés ou induisant un arrêt de sa traduction, ou si elle est localisée dans une séquence régulatrice de l'expression génique, induisant une variation de la quantité de transcrit. Les analyses d'associations pangénomiques que nous avons évoquées dans la section 2 visent à identifier les variations génétiques qui participent à la variabilité phénotypique. Seules 9 % des mutations mises en évidence par ces approches modifient une séquence protéique (Hindorff et al., 2009). De plus, 43 % des sites associés à une variabilité phénotypique sont localisés dans des séquences intergéniques qui sont apparues au cours de ces dernières années comme participant à la régulation de l'expression des gènes (Schaub et al., 2012). La majorité des sites identifiés par GWAS semble donc participer au phénotype *via* un contrôle de la transcription.

Les niveaux d'expression génique peuvent servir de phénotype d'étude intermédiaire entre un trait complexe observé au niveau de l'organisme et les variations génétiques (Jansen and Nap, 2001). De plus, les variations portant sur la régulation de la quantité des transcrits sont estimés comme jouant un rôle important dans l'adaptation locale des populations à leur environnement (Fraser, 2013). Caractériser la diversité génétique présente dans les régions régulatrices et tenter de la corréler à l'expression des séquences codantes revient donc à identifier des variations génétiques qui sont impliquées dans la variabilité de phénotypes plus complexes et qui peuvent être le fruit d'une adaptation de la population à un environnement (Gaffney, 2013). Un marqueur génétique de la quantité d'un transcrit est appelé eQTL de l'anglais *expression Quantitative Trait Loci*. Les sites identifiés par GWAS ont significativement plus de chance d'être associés à des différences d'expression génétique que ce qui est attendu par hasard (Nica et al., 2010; Nicolae et al., 2010). La cartographie d'eQTL se révèle donc complémentaire des GWAS dans le but d'identifier les causes génétiques des variabilités de phénotypes complexes.

Les cartographies d'eQTL ont l'avantage de requérir une cohorte moins importante que les GWAS (~50-200 individus non apparentés) pour avoir un pouvoir de détection de l'association suffisant. Leur efficacité à établir un lien entre génotype et phénotype moléculaire dépend aussi du nombre de transcrits dont l'expression est mesurée. Le développement des puces d'expression et, plus récemment, des techniques de séquençage des

ARN, a donc été une étape critique pour leur réalisation et permettent aujourd'hui d'obtenir des résultats particulièrement détaillés et non biaisés lors de l'utilisation des techniques de séquençage. De la même manière, l'amélioration constante des techniques de séquençage du génome permet d'identifier de plus en plus de variations génétiques, augmentant la résolution de la cartographie des eQTL.

### 4.1.2. Organisation génomique des eQTL

Au cours des dernières années, de nombreuses cartographies d'eQTL ont été réalisées dans différentes populations humaines (par exemple Lappalainen et al., 2013; Pickrell et al., 2010; Stranger et al., 2012)

Les eQTL peuvent être situés à différentes distances des gènes. Ils peuvent être proximaux et agir en *cis* ou être distaux et agir en *trans*. Les signaux d'associations les plus forts ont souvent été localisés à proximité du site d'initiation de la transcription (Montgomery and Dermitzakis, 2011). Ces résultats suggèrent que les régions génomiques proches des gènes tels que le promoteur et les sites de fixation des facteurs de transcription ont un effet particulièrement fort sur l'expression des gènes. Cependant, l'étendue de la cartographie des *trans*-eQTL requiert des corrections contraignantes, ce qui pourrait laisser supposer une potentielle perte de signal expliquée par la nécessité de la maitrise du nombre de faux résultats positifs du test. Par exemple, la recherche de *trans*-eQTL dans le cadre de l'immunité n'a permis de détecter que deux sites, chacun contrôlant l'expression de nombreux transcrits, qui seraient d'importants régulateurs de la quantité de transcrits dans les cellules (Fairfax et al., 2014, 2012). De plus, alors que les *cis*-eQTL ont un effet fort sur l'expression du gène dont ils sont proches, il a été proposé que les *trans*-eQTL ont, eux, un effet plus modeste sur l'abondance des ARNm, ce qui laisse supposer un rôle biologique plus important que ce qui est aujourd'hui démontré pour cette classe d'association génotype-phénotype (Gilad et al., 2008; Majewski and Pastinen, 2011; Montgomery and Dermitzakis, 2011).

L'influence de certains *trans*-eQTL sur un grand nombre de transcrits peut mener à l'identification de voies de signalisation grâce à une approche non biaisée par le choix de candidats d'étude. Ainsi, il a récemment été démontré qu'un locus lié à l'expression de nombreux gènes impliqués dans la réponse à l'interféron dans le cadre du lupus érythémateux disséminé était aussi corrélé à l'expression du facteur de transcription IKZF1 (Westra et al.,

2013). Les gènes régulés en *trans* par ce SNP se sont révélés enrichis en séquences servant à la fixation d'IKZF1, permettant de lever le voile sur le rôle de ce facteur de transcription dans la maladie systémique auto-immune étudiée. De nombreux autres *trans*-eQTL ont été identifiés dans la région génomique contenant les gènes codant pour le CMH (Fairfax et al., 2014, 2012; Fehrmann et al., 2011; Westra et al., 2013). Ces locus ont été associés à des différences d'expression du CMH-II dans des contextes particuliers (Fairfax et al., 2014), soulignant l'importance de l'effet de l'environnement sur ces associations.

### 4.1.3. eQTL et identification d'interactions entre gènes et environnement

Les phénotypes complexes résultent de l'interaction de nombreux facteurs génétiques et environnementaux. Cette complexité rend une étude globale impossible, ou tout du moins peu puissante. En revanche, réduire l'analyse à des approches cellulaires permet de mieux comprendre l'interaction entre les facteurs extérieurs et les variations génétiques impliquées dans les phénotypes non mendéliens (Dermitzakis, 2012; Idaghdour and Awadalla, 2012; Maranville et al., 2012). La cartographie d'eQTL permet d'atteindre ce niveau de détails. Des études récentes ayant utilisé différentes approches ont ainsi montré qu'une même variation génétique pouvait avoir des conséquences différentes sur l'expression de certains gènes en fonction de l'état d'activation de la cellule (Figure 10, Barreiro et al., 2012; Dombroski et al., 2010; Fairfax et al., 2014; Gargalovic et al., 2006; Lee et al., 2014; Maranville et al., 2011; Romanoski et al., 2010; Smirnov et al., 2009; Smith and Kruglyak, 2008), identifiant des eQTL de réponse. Les travaux réalisés chez l'Homme ont identifié des eQTL de réponse corrélés à l'expression de gènes suite à la stimulation de différents types cellulaires par des facteurs variés. Deux études ont à ce jour utilisé une stimulation des cellules par un agent infectieux, permettant d'identifier des facteurs génétiques impliqués dans les différences interindividuelles de réponses immunitaires (Barreiro et al., 2012; Lee et al., 2014). De telles stratégies d'étude permettent donc d'avoir une vue plus détaillée du contrôle génétique de l'expression des gènes, qui peut se faire dans un contexte environnemental particulier uniquement ou qui peut varier en fonction des facteurs externes à la cellule. Les SNP contrôlant l'abondance de transcrits et dont l'effet dépend de facteurs environnementaux se sont montré enrichis en locus liés à un risque de maladies infectieuses dans deux analyses (Barreiro et al., 2012; Fairfax et al., 2014). Cela traduit une fois de plus l'importance de la régulation de l'expression des gènes dans le devenir de l'infection et la susceptibilité aux

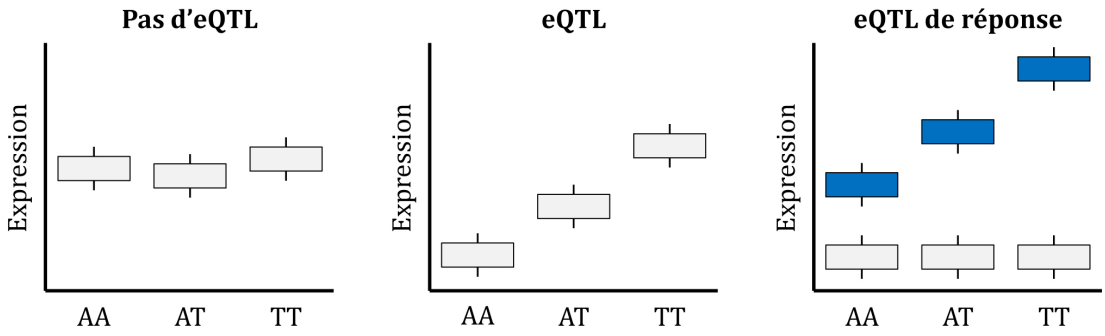



**Figure 10. Régulation génétique de l'expression génique.**
Pas d'eQTL : aucune association n'est retrouvée entre le génotype à un site donné et l'expression du gène. eQTL : l'expression du gène est corrélée au génotype au SNP considéré. eQTL de réponse : alors qu'aucune association entre génotype et expression génique n'est observée dans les cellules non stimulées (gris), on observe une corrélation entre génotype et phénotype suite à l'activation des cellules par un facteur extérieur (par exemple l'infection, en bleu).

maladies infectieuses. Si nous avons pour l'instant discuté du contrôle génétique de cette expression, il parait aussi d'intérêt de considérer les transcrits impliqués dans la régulation de la quantité des ARNm.

## 4.2. Les miARN, régulateurs de l'expression des gènes

Ces dernières années, les progrès faits dans l'identification et la quantification des ARN ont abouti à la découverte de nombreux transcrits non-codants régulant l'expression des gènes. Les plus célèbres de ces ARN sont sans conteste les microARN (miARN) qui sont les plus abondants des ARN régulateurs chez les eucaryotes supérieurs. Ces séquences sont très courtes (~22nt) et participent à la répression de la production des protéines par un mécanisme de régulation post-transcriptionnelle.

### 4.2.1. Identification des miARN

Le premier miARN a été identifié chez des larves du nématode *Caenorhabditis elegans* (Lee et al., 1993). Lors du développement larvaire, le gène *lin-14* joue un rôle capital pendant le stade L1 mais son expression doit être réprimée pour permettre la poursuite du cycle. Le gène

*lin-4* code pour un inhibiteur spécifique de l'expression de *lin-14* au cours de la transition L1-L2. Ce gène de 693 pb ne possède pas de phase ouverte de lecture. Le transcrit ainsi produit est non-codant. Il existe deux formes de *lin-4* dans la cellule : une longue, faisant 61 nt, et une courte qui ne fait que 21 nucléotides. *Lin-4L* présente une structure secondaire particulière, en forme d'épingle à cheveux, alors que *lin-4S* contient des séquences complémentaires de sept motifs nucléotidiques présents dans le 3' UTR de *lin-14*, suggérant que la fixation de cette courte séquence sur le transcrit de *lin-14* induit une répression de l'expression de la protéine correspondante. Il a ensuite fallu attendre sept ans pour qu'un autre gène codant pour un petit transcrit ayant des propriétés similaires soit identifié. En 2000, une étude a montré que *let-7* dont l'expression est contrôlée au cours du développement du nématode  code pour un petit ARN de 21 nt qui inhibe l'accumulation des protéines LIN-41 et LIN-42 au cours de la transition entre le stade larvaire L4 et le stade adulte (Reinhart et al., 2000; Slack et al., 2000). Le modèle de régulation de l'expression des gènes ainsi mis en évidence a été généralisé et des séquences homologues à *let-7* ont été identifiées chez toutes les espèces d'invertébrés et de vertébrés à symétrie bilatérale étudiées. Cette forte conservation suggère un rôle fonctionnel particulièrement important de cet ARN non-codant.

Par la suite, les premiers balayages systématiques ont été réalisés pour identifier d'autres miARN. Malgré les limites techniques du séquençage de type Sanger, des centaines de ces séquences ont été mises en évidence chez les vers, les mouches et les mammifères, montrant que ces petites séquences régulatrices étaient beaucoup plus présentes que ce qui était précédemment estimé (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001). Afin de contourner les limites techniques imposées par le clonage et le séquençage de séquences, des outils informatiques visant à scanner les génomes dans le but de trouver des séquences codantes dont les transcrits présentent les caractéristiques des miARN ont été largement développées. Ces méthodes s'appuient sur plusieurs principes, notamment sur la conservation de séquence et la présence d'un transcrit précurseur présentant une structure secondaire en forme d'épingle à cheveux (Figure 11, (Grad et al., 2003; Lai et al., 2003). Cependant, ces analyses informatiques ont montré leurs limites avec l'identification de très nombreuses régions génomiques qui correspondent aux critères recherchés mais ne sont malheureusement pas transcrites. Le développement des techniques de séquençage à très haut débit permettant notamment de cibler les ARN ont modifié la donne. En effet, ces analyses

ont permis de mesurer l'abondance de petits transcrits de manière non biaisée dont les propriétés inférées par études informatiques se rapprochent de celles des miARN.

### 4.2.2. Base de données et nomenclature

L'identification d'un très grand nombre de séquences correspondant à des miARN a rapidement mené à la nécessité d'établir une nomenclature unifiée ainsi qu'une base de données recensant l'ensemble des transcrits ainsi mis en évidence. Ces besoins ont mené à la création de miRBase (Griffiths-Jones et al., 2008) qui est un répertoire de toutes les séquences de miARN décrites dans la littérature (1881 séquences chez l'Homme dans la version la plus récente) et qui établit des critères à suivre pour identifier, annoter et classifier les miARN (Ambros et al., 2003). Concernant la nomenclature, le préfixe « miR- » est utilisé pour toutes les séquences. Il est suivi d'un nombre et parfois d'une lettre lorsque plusieurs gènes ont des séquences excessivement proches (miR-29a et miR-29b par exemple). Un chiffre peut ensuite être utilisé si deux séquences codantes localisées dans des régions génomiques distinctes mènent à la production du même transcrit (miR-29b-1 et miR-29b-2 par exemple). Finalement, un miARN peut provenir d'un bras ou l'autre du précurseur. Une

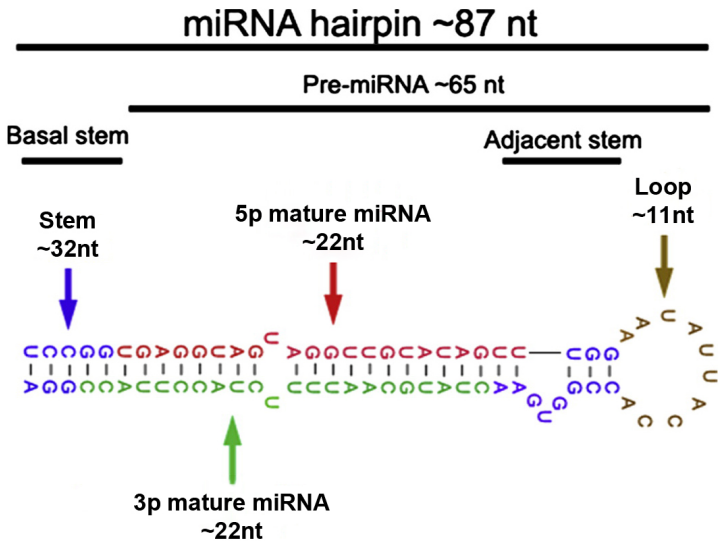



**Figure 11. Structure des précurseurs de miARN.**
Adapté de Quach et al. (2009). La structure secondaire caractéristique en forme d'épingle à cheveux contient le précurseur du miARN ainsi que les séquences avoisinantes qui forment la base de la tige. Le pre-miARN est composé d'un duplex imparfait entre les deux miARN matures 5p et 3p, de la région adjacente de la tige et d'une boucle terminale. La désignation des deux miARN se fait en fonction de la proximité des extrémités de la tige (« 5p » ou « 3p ») et remplace les annotations antérieures de formes mature (miR) et complémentaire (miR*).

ancienne nomenclature parlait de miARN mature et de la forme étoile, qui référait au miARN du bras opposé. Cette distinction se basait sur l'abondance relative de chacun des deux transcrits, la forme étoile étant moins présente que la forme dominante. Cependant, ces notions de forme majoritaire ont au fur et à mesure été mises en question et dans un souci d'harmonisation il convient maintenant de parler de formes « 3p » ou « 5p » (miR-29-b1-3p par exemple).

### 4.2.3. Biogénèse des miARN

#### 4.2.3.1. Voie canonique

Le contexte génomique dans lequel se situent les miARN peut être très variable. Les miroARN canoniques sont transcrits par l'ARN Polymérase II puisqu'ils sont localisés dans des régions intergéniques ou dans les introns d'autres gènes (Lee et al., 2004). Un transcrit primaire (pri-miARN) qui peut contenir une (mono-) ou plusieurs miARN en forme d'épingles à cheveux (poly- cicstronique) est alors synthétisé. Si certains miARN transcrits en même temps partagent des similarités de séquence ou de fonctions (miR-212-3p et miR-132-3p sont par exemple prédits comme ciblant les mêmes ARNm chez l'Homme, Wanet et al., 2012), cette propriété n'est pas une généralité.

Les promoteurs utilisés pour l'expression des miARN ne sont pas forcément connus. C'est d'autant plus le cas pour les miARN intergéniques. Cependant, des études d'expression de ces gènes particuliers ont montré que les quantités de transcrits de miARN localisés à moins de 50kb les uns des autres sont corrélées, suggérant qu'ils sont transcrits ensemble sous forme polycistronique (Baskerville and Bartel, 2005). Près de la moitié des miARN sont localisés dans des introns. L'expression d'une partie d'entre eux est corrélée avec celle des gènes dans lesquels ils sont situés (Baskerville and Bartel, 2005; Rodriguez et al., 2004), illustrant une co-transcription. Cependant, nombre d'études ont remis ce modèle en question et suggèrent plutôt que les miARN introniques ont leur propre promoteur (Martinez et al., 2008; Parts et al., 2012).

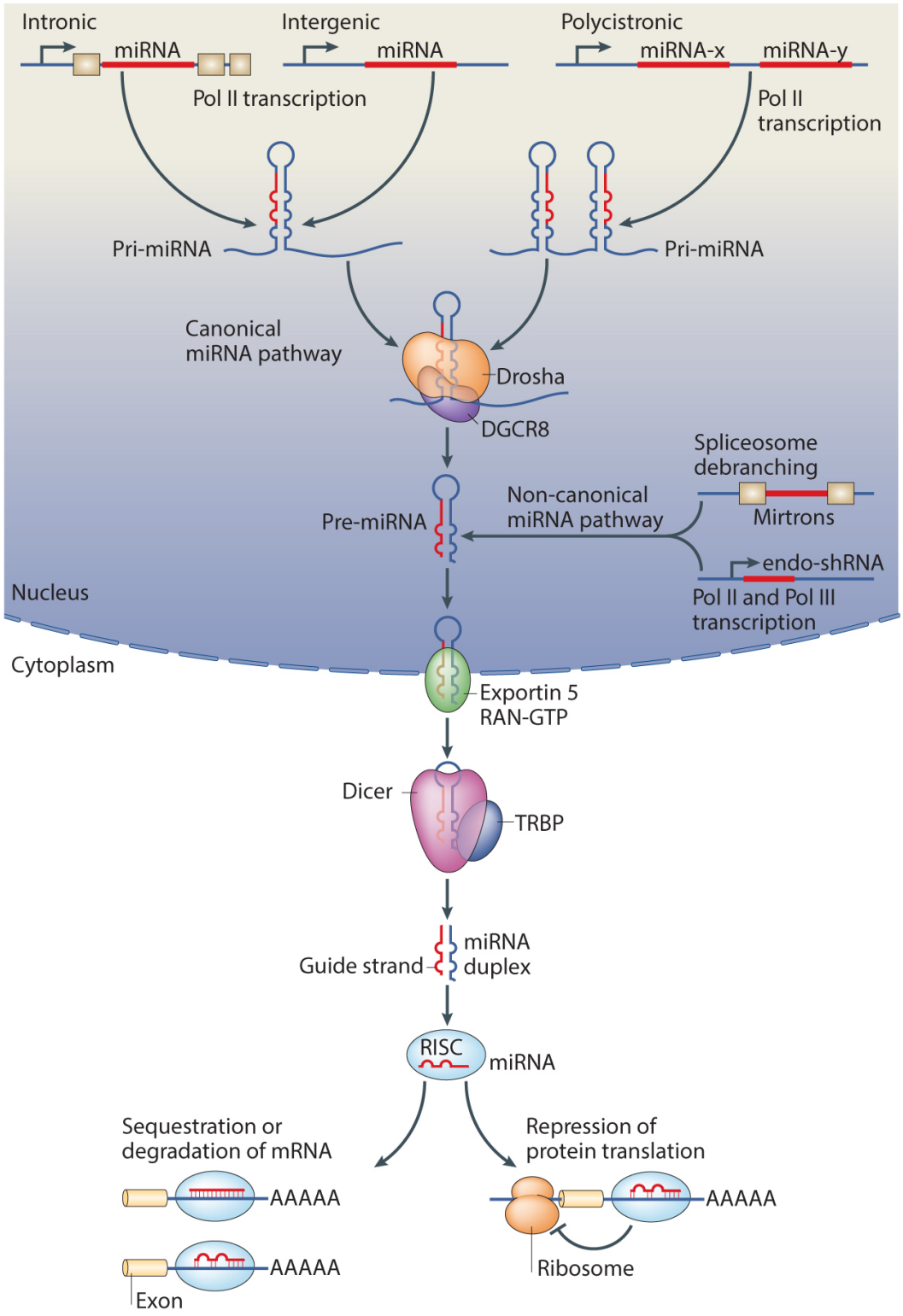



**Figure 12. Synthèse des miARN et régulation post-transcriptionnelle des ARNm.**
Tiré de Rottiers and Näär (2012). Le précurseur de miARN est produit dans le noyau par voie dépendant ou non de Drosha. Une fois exporté dans le cytoplasme, il est clivé par Dicer. Le duplex ainsi obtenu contient deux miARN. Un des brins du complexe sera préférentiellement chargé dans le complexe miRISC et guidera la régulation de la quantité des ARNm cibles par inhibition de la traduction et/ou dégradation du transcrit.

Le pri-miARN ainsi synthétisé va subir deux clivages successifs par des ribonucéases de type III. La première à agir est DROSHA, qui reconnaît les ARN double-brins nucléaires et qui clive le pri-miARN en un produit d'environ 70-80 nt appelé « précurseur » ou pre-miARN (Figure 12, Lee et al., 2003). Cet intermédiaire est replié en une structure secondaire en tige-boucle dont l'extrémité 5' porte un phosphate alors que les deux derniers nucléotides de l'extrémité 3' sont libres, non appariés à des bases complémentaires (Figure 11). L'action spécifique de DROSHA nécessite l'intervention d'un autre facteur moléculaire, DGCR8, qui permet d'obtenir un clivage à une position stable des pri-miARN malgré la diversité de formes qu'ils peuvent prendre (Han et al., 2006). Le pre-miARN est ensuite exporté du noyau vers le cytoplasme par la protéine XPO5 (Exportin-5) où il est reconnu et pris en charge par DICER (Figure 12). Cette endoribunucléase clive la boucle du pre-miARN pour aboutir à l'obtention d'un complexe formé par deux séquences de 22 nt imparfaitement complémentaires, le miARN 3p et le miARN 5p (Hutvágner et al., 2001; Ketting et al., 2001; Park et al., 2011).

#### 4.2.3.2. Voie de synthèse alternative

De plus en plus de miARN sont décrits comme étant produits d'une manière différente de celle que nous venons d'évoquer. Ces miARN sont issus de la transcription de certains gènes ne codant pas pour des protéines et des éléments transposables. L'une de leur particularité est qu'ils n'ont pas nécessairement besoin d'être clivés par DROSHA. Même si des preuves de la fonctionnalité de ces séquences s'accumulent, leur mise en évidence et surtout leur validation est difficile, ce qui explique certainement en partie pourquoi cette voie de synthèse alternative semble largement minoritaire. Une autre voie plus utilisée est celle suivie par les miARN qui sont localisés dans des introns qui prennent une structure secondaire de tige-boucle lors de l'épissage des exons, structure qui est semblable à celle d'un pre-miARN. Ces transcrits n'ont donc pas besoin d'être clivés par DROSHA (Figure 12). Ils sont appelés mirtrons et ont été largement décrits dans le monde vivant (Berezikov et al., 2007; Okamura et al., 2007; Ruby et al., 2007).

### 4.2.4. Répression de l'expression des gènes

Mis à part un nombre particulièrement restreint d'exceptions (Vasudevan et al., 2007), les miARN répriment l'expression génique avec des effets modestes sur la quantité finale de protéine (Bartel and Chen, 2004). Ces transcrits servent donc à la modulation fine, précise de l'expression des gènes. Une fois le second clivage effectué par DICER, le duplexe de miARN est séparé et un des brins est chargé dans le *RNA-induced silencing complex* (le complexe RISC) composé de plusieurs acteurs protéiques dont un des membres de la famille des protéines Argonaut (AGO) (Figure 12, Peters and Meister, 2007). Chez l'Homme, il semblerait qu'AGO2 soit le membre de cette famille protéique le plus utilisé dans la formation du complexe RISC (Meister et al., 2004). Cependant, les autres protéines Argonaute sont aussi associées à une répression de la traduction médiée par les miARN. La manière dont sont discriminés les miARN simple brin lors du chargement du complexe RISC reste relativement floue. Il a dans un premier temps été proposé qu'une différence d'expression pourrait être la cause principale de cette distinction (Schwarz et al., 2003). Le brin non chargé pourrait être moins stable, donc rapidement dégradé, ce qui explique qu'il ne soit pas retrouvé dans les complexes miARN-RISC (Khvorova et al., 2003; Schwarz et al., 2003). Cependant, des données montrant que les deux bras peuvent être retrouvés en proportions stœchiométriques remettent cette hypothèse en cause (Lagos-Quintana et al., 2002). Les techniques de séquençage actuelles permettent de mieux apprécier les quantités relatives de chacun des bras du complexe miARN et ont notamment permis de montrer que la forme dominante pouvait changer en fonction des tissus ou des stades du développement (Chang et al., 2012; Chiang et al., 2010; Cloonan et al., 2011; Li et al., 2012). Qui plus est, cela traduit une dynamique de l'expression et de la fonctionnalité des miARN qui était jusque-là insoupçonnée.

La reconnaissance de l'ARNm ciblé se fait par complémentarité de séquence entre les nucléotides 2 à 8 de la partie 5' du miARN, appelée *seed sequence*, et des séquences de l'ARNm ciblé. Les sites de fixation des miARN se trouvent principalement dans la partie 3' des transcrits cibles mais il a récemment été montré que d'autres régions localisées dans la partie 5' voire dans les régions codantes pouvaient servir à la régulation de la traduction des ARNm par les miARN même si elles semblent moins efficaces dans ces deux derniers cas (Grimson et al., 2007; Lytle et al., 2007; Schnall-Levin et al., 2010).

Si la complémentarité entre le miARN et le site de fixation est parfaite, l'activité endonucléase de la protéine AGO engendre le clivage et la dégradation de l'ARNm ciblé. C'est le mécanisme mis en jeu lors de la régulation de l'expression de HOXB8 par miR-196 (Yekta et al., 2004). Cependant, les cas de complémentarité parfaite sont relativement rares dans le monde animal. Malgré tout, de nombreuses études ont rapporté une corrélation négative entre l'expression des miARN et de leurs ARNm cibles (voir par exemple Wang and Li, 2009). De plus, les nombreuses études ayant utilisé des outils moléculaires pour perturber artificiellement la quantité d'un miARN donné dans les cellules montrent des conséquences sur l'abondance de transcrits présentant des sites de fixation de ce miARN particulier (Krützfeldt et al., 2005; Lim et al., 2005). Il semblerait donc que la régulation post-transcritpionnelle de l'expression des gènes exercée par cette classe de petits ARN soit le principal mécanisme d'action chez les mammifères (Baek et al., 2008; Selbach et al., 2008). Plutôt qu'un clivage par AGO-2, la déstabilisation de l'ARNm mise en jeu en cas d'appariement imparfait impliquerait des mécanismes de déadénylation, de « décoiffage » 5' et d'une dégradation 5' -> 3' qui sont à l'origine de la voie de recyclage des transcrits (Eulalio et al., 2009a; Huntzinger and Izaurralde, 2011).

Un autre mécanisme de régulation de la production des protéines par les miARN est une répression de la traduction. Cette notion a été formulée très tôt puisque l'expression de *lin-4*, dont nous avons parlé plus tôt, n'influence pas la quantité de transcrits *lin-14* et *lin-18* (Olsen and Ambros, 1999). Cette inhibition peut se faire au niveau de l'initiation, l'élongation ou de la terminaison de la traduction. Il est aussi possible d'avoir une inhibition co-traductionnelle (Filipowicz et al., 2008). Le premier processus semble être le plus répandu et impliquerait la protéine GW182 (Eulalio et al., 2009b, 2008).

### 4.2.5. Les miARN et la réponse immunitaire

Les miARN sont impliqués dans de nombreux processus cellulaires, du développement à l'apoptose en passant par la différenciation cellulaire et la réponse à des stimulations externes (Ambros, 2004). La forte conservation de la séquence participant à la reconnaissance des ARNm montre que leur rôle de régulateurs post-transcriptionnels de l'expression des gènes est capital pour la cellule (Quach et al., 2009).

Parmi les fonctions associées aux miARN, leur implication dans les réponses mises en place suite à l'infection par des pathogènes est de mieux en mieux décrite (Lodish et al., 2008; O'Connell et al., 2012, 2010). Ces transcrits ont été démontrés comme régulant l'expression de nombreux gènes au cours de la réponse immunitaire innée ainsi qu'adaptative et leur rôle dans le développement des cellules immunitaires et la modulation de l'intensité des réponses inflammatoires a été particulièrement étudié au cours de ces dernières années (Chen et al., 1995; Johnnidis et al., 2008; O'Connell et al., 2012). Pour se restreindre à l'immunité innée, de nombreux miARN ont été décrits dans le cadre de la régulation de diverses voies de signalisation, que ce soit directement au niveau de l'expression des récepteurs tels que les RLR, les TLR (Figure 13) ou les NLR (He et al., 2014; Li and Shi,

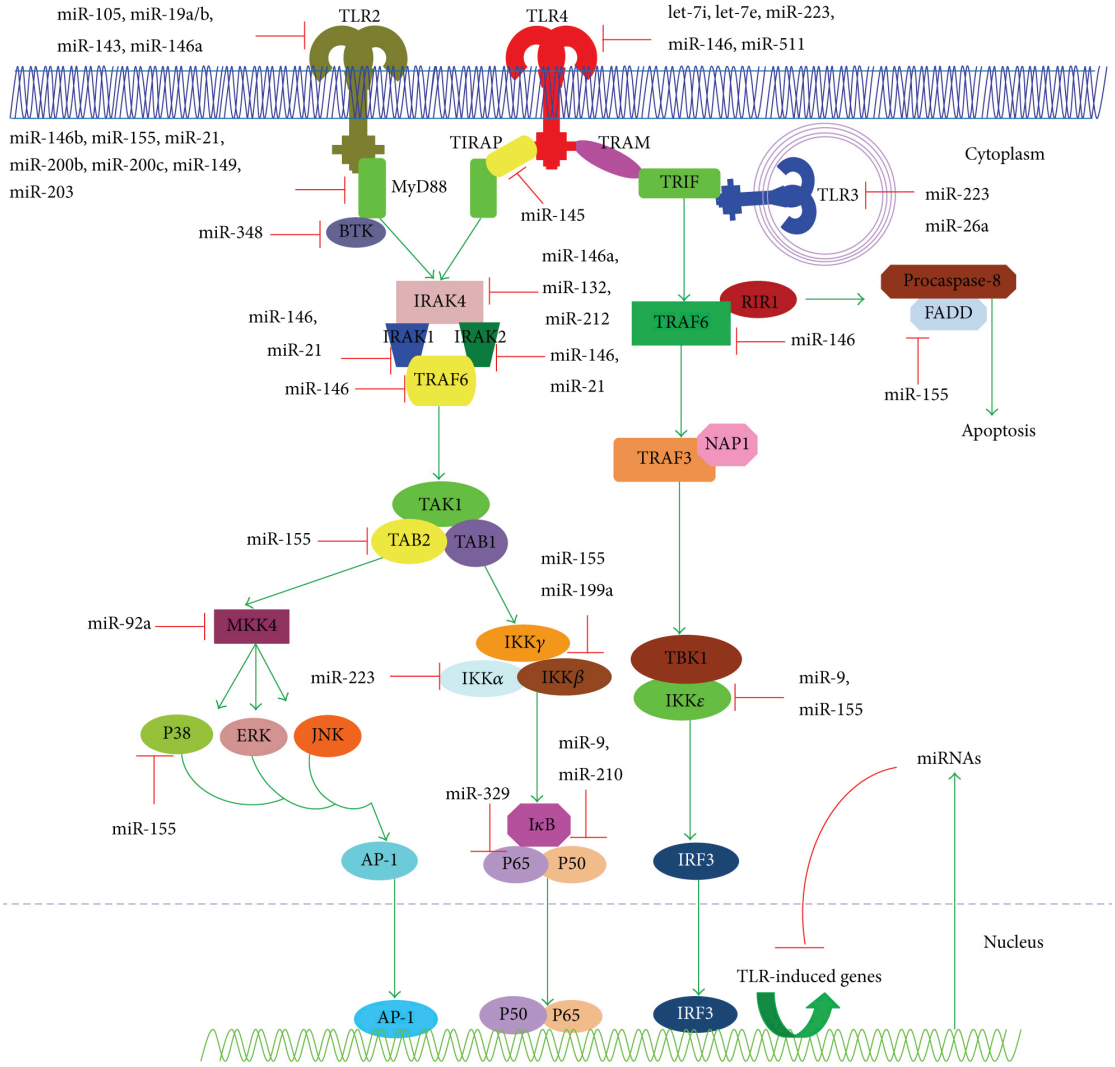



**Figure 13. Régulation de la voie de signalisation impliquant les TLR par les miARN.**
Tiré de He et al. (2014). Les miARN régulent la voie TLR à différents niveaux, des récepteurs aux molécules effectrices.

2013; O'Neill et al., 2011), au niveau de l'expression des molécules de transduction du signal ou au niveau de la production d'effecteurs de la réponse immunitaire (Taganov et al., 2006).

La première étude portant sur l'expression des miARN dans le cadre d'une réponse de cellules immunitaires à une stimulation a ainsi montré l'induction de trois transcrits, miR-155, miR-146a et miR-132 dans les monocytes suite à leur activation par le lipopolysaccharide (LPS) (Taganov et al., 2006). Ces miARN sont non seulement induits par la voie de signalisation en aval de TLR4 mais sont aussi eux-mêmes impliqués dans la régulation de cette voie à différents niveaux. Par exemple, miR-155 inhibe l'expression de MYD88 qui est l'une des premières molécules activées suite à la stimulation des TLR localisés à la membrane cytoplasmique, ainsi que celle de TAB2, IKBKE et FADD qui sont en aval de la voie de signalisation mise en jeu (Figure 13).

D'autres données suggèrent aussi que les miARN peuvent être utilisés comme mécanisme de défense des cellules immunitaires. Par exemple, miR-32 se fixe sur la partie 3' de plusieurs ARNm produits par le rétrovirus PFV1 (Primate Foamy Virus type 1), participant à l'inhibition de la réplication de ce pathogène (Lecellier et al., 2005). De la même façon, plusieurs miARN dont l'expression est induite par l'IFNβ montrent une complémentarité de séquence presque parfaite avec des sites présents dans le génome du virus de l'hépatite C et participent à la réponse antivirale des cellules (Pedersen et al., 2007) alors que l'expression de miARN favorables à la réplication du virus est, elle, inhibée (Jopling, 2012; Jopling et al., 2005; Pedersen et al., 2007). L'ensemble de ces résultats montre que les miARN sont non seulement des produits de la réponse immunitaire mais participent activement, largement et de façon concertée avec d'autres processus cellulaires aux défenses mises en place suite à une stimulation par des agents pathogènes.

## 4.3. Différences de réponses immunitaires et tuberculose

### 4.3.1. Données épidémiologiques

#### 4.3.1.1. Données historiques

La tuberculose n'est pas une maladie ayant touché l'espèce humaine récemment. Les premiers cas de tuberculose qu'il a été possible d'identifier remontent à 7000 ans avant JC en Europe et au Moyen-Orient. Cette datation a été réalisée à partir de l'étude de squelettes du

Néolithique qui présentaient des signes de spondylodiscite causée par le bacille de la tuberculose, aussi appelé Mal de Pott, ainsi que par des rapports retrouvés sur des tablettes assyriennes (Haas and Haas, 1996; Hershkovitz et al., 2008). D'autres données suggèrent que les premières infections pourraient même pré-dater la révolution néolithique (Roberts et al., 2009). Cependant, les premières épidémies ont touché l'espèce humaine à partir du XVIème siècle et ont atteint leur paroxysme au début du XIXème siècle, favorisées par les conditions sanitaires déplorables, la surpopulation et la révolution industrielle (Haas and Haas, 1996). L'amélioration des conditions d'hygiène, la découverte de la souche BCG suivie de son utilisation comme vaccin dès les années 1920 (Calmette et al., 1925) et l'introduction des chimiothérapies comme traitement ont permis de réduire l'incidence de la maladie au cours du XXème siècle (Haas and Haas, 1996).

#### 4.3.1.2. Données actuelles

Malgré les progrès réalisés en termes de dépistage et de traitement, la tuberculose reste un problème de santé publique majeur à travers le monde. Le principal agent de la tuberculose, *Mycobacterium tuberculosis* (MTB), est responsable du développement d'une forme active de la maladie chez 9 millions d'individus par an et de la mort de près de 1,5 millions de personnes chaque année, ce qui en fait le deuxième agent infectieux le plus meurtrier derrière le VIH. 85 % des cas de tuberculose sont localisés en Afrique et en Asie. Les conditions sanitaires parfois désastreuses et la malnutrition font partie des facteurs environnementaux qui participent à la propagation de la maladie (Hargreaves et al., 2011; Lönnroth et al., 2009). Il existe une forte relation entre l'immunodéficience causée par l'infection par le VIH et la tuberculose. L'inertie entre ces deux maladies infectieuses est particulièrement préoccupante : au moins un tiers des personnes vivant avec le VIH sont infectées par MTB même s'ils ne présentent pas systématiquement des symptômes et près de 25 % des décès de patients infectés par le VIH sont dus à la tuberculose (World Health Organization, 2014). Les deux pathogènes accélèrent mutuellement leur progression et un individu immunodéprimé présente 29 à 31 fois plus de chances de développer une forme active et évolutive de la tuberculose qu'une personne non infectée par le VIH (World Health Organization, 2014). L'autre principal problème lié à la tuberculose est la recrudescence de souches multirésistantes et ultrarésistantes, qui, respectivement, ne sont pas sensibles aux principaux traitements médicamenteux prescrits et qui ne répondent qu'à un nombre encore plus restreint

de médicaments. 480 000 cas de tuberculose multirésistante ont ainsi été recensés en 2013 chez des patients présentant une forme active de tuberculose pulmonaire, dont 9 % étaient probablement des souches ultrarésistantes (World Health Organization, 2014).

### 4.3.2. Génétique des différences de susceptibilité à la tuberculose

De nombreux facteurs environnementaux vont influencer l'infection et/ou le développement de la maladie. Cependant, malgré une prévalence élevée de la tuberculose, seules 5 à 10 % des personnes infectées vont développer des symptômes cliniques signes d'une forme active de la maladie infectieuse (World Health Organization, 2014). De plus, le taux de mortalité lié à la tuberculose n'est pas le même en fonction de l'âge des patients. La courbe de mortalité a une forme particulière en U, traduisant une forte mortalité aux âges extrêmes qui est moindre aux âges intermédiaires. Cela peut être expliqué par les primo-infections et la réactivation de la maladie ou des infections secondaires (Alcaïs et al., 2010, 2005). De plus, cela reflète différents mécanismes de susceptibilité (Alcaïs et al., 2005). Une différence de susceptibilité très importante peut être observée entre diverses populations humaines (Hoeppner and Marciniuk, 2000; Sousa et al., 1997), expliquée par les exposition passées des populations à MTB. Enfin, des études réalisées sur des jumeaux ont montré dès la seconde moitié du XXème siècle que la susceptibilité à la tuberculose avait aussi des bases génétiques, héréditaires (Comstock, 1978). De nombreuses études ont par la suite identifié certains gènes participant à la réponse à l'infection par MTB et pouvant expliquer une partie des différences de susceptibilité interindividuelles observées (Fortin et al., 2007). Cette susceptibilité a été décrite chez l'homme comme pouvant être mendélienne ou complexe, impliquant des variants effets, pénétrances et fréquences alléliques très divers (Abel and Casanova, 2000; Alcaïs et al., 2005; Casanova and Abel, 2002).

#### 4.3.2.1. Susceptibilité mendélienne à la tuberculose

Comme nous l'avons précédemment évoqué au cours de cette introduction, les immunodéficiences primaires prédisposent les patients qui en sont atteints à certaines maladies infectieuses. On retrouve donc des IP associées à la tuberculose. Les individus présentant de tels troubles de la réponse immunitaire sont prédisposés aux infections par des souches mycobactériennes peu virulentes telles que la souche vaccinale BCG ainsi qu'à

d'autres pathogènes comme *Salmonella* et certains virus (Casanova and Abel, 2002; van de Vosse et al., 2013; Zhang et al., 2008). On parle alors de prédisposition mendélienne aux maladies mycobactériennes ou MSMD (pour *Mendelian Susceptibility to Mycobacterial Diseases*). Neuf gènes associés aux MSMD ont été identifiées : *NEMO* (Filipe-Santos et al., 2006) et *CYBB* (Bustamante et al., 2011) sont localisés sur le chromosome X et *IFNGR1* (Jouanguy et al., 1996; Newport et al., 1996), *IFNGR2* (Vogt et al., 2008, 2005), *STAT1* (Dupuis et al., 2001), *IL12B* (Altare et al., 1998b), *IL12RB1* (pour lequel plus de 70 mutations ont été trouvées) (Altare et al., 1998a; de Jong et al., 1998), *IRF8* (Hambleton et al., 2011) et *ISG15* (Bogunovic et al., 2012) sont localisés sur des autosomes (pour une revue des données immunologiques, génétiques et cliniques des MSMD, se référer à Bustamante et al., 2014). On estime que 45 % des jeunes patients développant une tuberculose miliaire (plus connue sous le nom de « tuberculose disséminée ») présentent une MSMD. On peut noter de manière particulièrement intéressante que l'ensemble des protéines codées par ces gènes sont impliquées dans la voie de signalisation IL-12-IL-23-IFNγ qui est particulièrement importante pour la communication entre les monocytes, macrophages ou cellules dendritiques d'un côté et lymphocytes T ou cellules NK de l'autre, menant à la différentiation des lymphocytes T naïfs en lymphocytes auxiliaires de type 1 (Figure 14, Chapman and Hill, 2012). L'implication de ces gènes dans la MSMD montre l'importance de cette voie de signalisation dans la réponse à l'infection par les mycobactéries. Des mutations dans les gènes *IL12RB1* et *IFNGR1* ont aussi été identifiées dans des cas de tuberculose à l'âge adulte (Alcaïs et al., 2005; Tabarsi et al., 2011), suggérant que les variations génétiques peuvent parfois avoir une pénétrance incomplète et participer à une susceptibilité complexe.

#### 4.3.2.2. Susceptibilité non mendélienne à la tuberculose

Les approches de liaison pangénomique ou d'association concentrée sur des gènes candidats ont été les principales utilisées pour mettre en évidence les facteurs génétiques sous-jacents à la susceptibilité complexe à la tuberculose. Les résultats des études de liaison se sont révélés assez limités, ne permettant d'identifier qu'un faible nombre de sites (Möller and Hoal, 2010). Parmi les régions trouvées, un locus situé en 8q12-q13 fut par la suite étudié plus en détail. Il a ainsi été possible d'associer un variant génétique du gène *CYP7A1* à la susceptibilité à la tuberculose dans la population marocaine (Baghdadi et al., 2006; Qrafli et al., 2014). Des approches gènes-candidats ont aussi été menées et ont mis en évidence un

grand nombre de gènes impliqués dans la susceptibilité ou la résistance à la tuberculose (Fortin et al., 2007; Möller and Hoal, 2010). La première association ainsi établie, impliquant le gène *NRAMP1* dans des populations d'Afrique de l'Ouest, a par la suite été répliquée et étudiée à de nombreuses reprises (Bellamy et al., 1998; Möller and Hoal, 2010). D'autres associations ont pu être faites avec des gènes dont les protéines ont des fonctions moléculaires vairées, tels que les récepteurs CD209, TLR8 et le récepteur à la vitamine D ainsi que la molécule de transduction du signal STAT4 (Barreiro et al., 2006; Davila et al., 2008; Sabri et al., 2014; Wilkinson et al., 2000).

Une autre méthode d'investigation des causes génétiques de la susceptibilité à une maladie infectieuse est de faire une GWAS. La première étude de ce type réalisée sur des populations Africaines n'a pu identifier que des associations faibles. Certains des locus identifiés au cours de cette étude se trouvent dans la région contenant les gènes codant pour le HLA-DQ, un récepteur CMH de classe II (Thye et al., 2010). La meilleure association

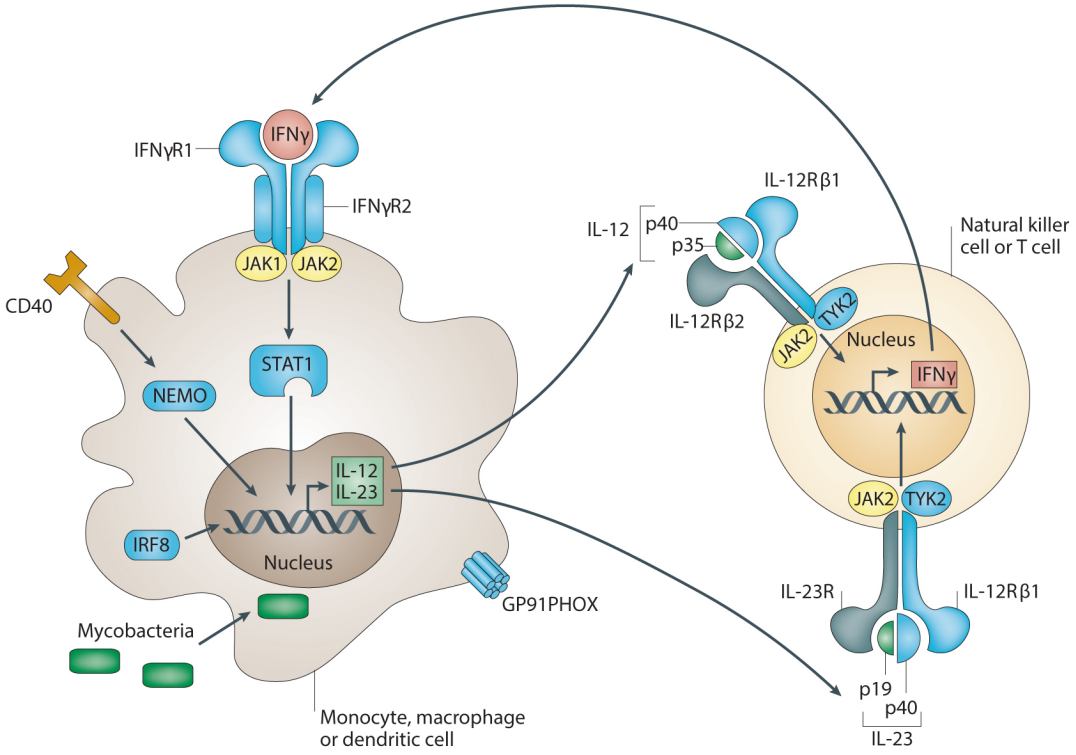



**Figure 14. Implication de la voie IL12-IL23-IFNγ dans les MSMD.**
Tiré de Chapman and Hill (2012). Des mutations dans des gènes participant à la synthèse d'IFNγ ou dans la réponse à cet effecteur ont été associées à des MSDM. Ces gènes sont indiqués en bleu clair. Il est important de noter que les variations localisées dans *TYK2* ont été associées à des susceptibilités à un large spectre de maladies infectieuses, dont les infections par les mycobactéries.

obtenue au cours de cette étude se révéla être une région pauvre en gène (en 18q11.2), montrant les limites de ce type d'approche. Par la suite, une autre région génomique fut associée à la susceptibilité à la tuberculose (Thye et al., 2012). Cependant, cette étude ne permit pas non plus d'identifier un gène dont les fonctions sont liées à la réponse de l'hôte à l'infection par les mycobactéries. Ce constat ainsi que la faible reproductibilité des résultats de GWAS (Chimusa et al., 2014) montrent les limites de ce type d'approche : d'une part, la contribution génétique dans la susceptibilité à la tuberculose semble varier en fonction des populations étudiées, d'autre part, elles n'ont pour l'instant permis de formellement identifier un mécanisme moléculaire à l'origine des différences de susceptibilité à cette maladie infectieuse que dans de très rares cas. Très récemment, une étude d'association pangénomique a associé des variations génétiques présentes dans les introns du gène *ASAP1* à la prédisposition à l'infection par MTB (Curtis et al., 2015). Un de ces SNP, rs10956514, induit une réduction de l'expression du gène dans les cellules dendritiques suite à l'infection par MTB. Des études *ex vivo* ont révélé que les cellules dendritiques n'exprimant pas *ASAP1* présentaient des défauts de dégradation de la matrice extracellulaire couplés à des défauts de migration, proposant de façon particulièrement robuste un processus cellulaire potentiellement associé à la prédisposition à la tuberculose.

#### 4.3.2.3. Approche transcriptomique de la susceptibilité à la tuberculose

Comme nous venons de l'évoquer, les études pangénomiques ont permis d'identifier des variants ségrégant dans la population associés à des phénotypes complexes mais ne fournissent que peu d'informations sur les mécanismes causaux. De plus en plus de données suggèrent que les variations non-codantes jouent un rôle important dans l'obtention d'un phénotype final *via* la régulation de l'expression des gènes (Dermitzakis, 2008). Par exemple, certains polymorphismes identifiés par analyses de gènes candidats sont localisés dans des promoteurs, comme dans le cas de *MCP1* (Flores-Villanueva et al., 2005; Thye et al., 2009), suggérant que des variations de l'expression de gènes pourrait mener à une meilleure appréhension des différences de réponse à l'infection par MTB. Une étude se basant sur cette hypothèse à récemment été conduite dans le but d'identifier les différences interindividuelles de réponses transcriptionnelles des cellules dendritiques suite à l'infection par MTB et d'identifier les facteurs génétiques qui sont à la base de ces dissemblances. Cette analyse a identifié 198 locus associés à des différences d'abondance de transcrits suite à la stimulation

des cellules dendritiques par le pathogène, locus qui se sont révélés enrichis en SNP précédemment liés à des différences de susceptibilité à la tuberculose par GWAS (Barreiro et al., 2012). Cette analyse a permis de proposer un nouveau gène comme impliqué dans la réponse à l'infection par MTB et de démontrer la puissance de ce type d'approche pour mieux comprendre les différences interindividuelles de susceptibilité aux maladies infectieuses. La question de la régulation de l'expression génique et du contrôle génétique d'une telle régulation reste cependant en suspens.

# Objectifs de la thèse

Ma thèse s'articule autour de deux axes principaux avec les objectifs suivants :

1) Étudier les signatures moléculaires des pressions sélectives exercées sur les gènes impliqués dans l'immunité innée chez l'Homme :

   (i) Mettre en évidence l'occurrence et l'intensité de la sélection purificatrice sur les gènes de l'immunité innée depuis la divergence entre l'Homme et le Chimpanzé,
   (ii) Identifier les gènes impliqués dans la réponse immunitaire immédiate présentant des signes d'une participation à l'adaptation des populations humaines à leur environnement,
   (iii) Dater l'occurrence des évènements de sélection positive dans ces séquences codantes,
   (iv) Estimer la part de la diversité génétique retrouvée dans les gènes de l'immunité innée qui résulte d'un flux de gènes provenant de l'Homme de Néandertal.

2) Caractériser la régulation de la réponse des cellules à l'infection exercée par les miARN ainsi que le contrôle génétique de l'expression de ces transcrits :

   (i) Caractériser la variabilité de réponse transcriptionnelle portant sur les miARN au cours de l'infection au sein de la population,
   (ii) Identifier les facteurs génétiques associés à l'expression de ces gènes au cours de la stimulation infectieuse,
   (iii) Estimer les effets de l'infection sur les réseaux de régulation de la quantité des ARNm par les miARN,
   (iv) Caractériser le rôle des miARN dans la régulation des réponses transcriptionnelles et protéiques des cellules dendritiques suite à l'infection en utilisant miR-29 comme modèle d'étude.

# Résultats

# 1. Evolution des gènes de l'immunité innée chez l'Homme

## 1.1. Contexte

Les pathogènes ont été une menace invisible persistante pour l'Homme au cours de son évolution. L'immunité innée est la réponse mise en place par l'organisme immédiatement après l'infection. Elle consiste en une reconnaissance de motifs divers relativement conservés dans de larges familles de micro-organismes qui engendre une réponse cellulaire semi-spécifique (Medzhitov, 2001). Les gènes codant pour des protéines impliquées dans ce processus constituent donc un excellent modèle d'étude permettant de mieux comprendre la manière dont les pathogènes ont participé au modelage du génome humain. Comme nous l'avons vu dans l'introduction, de nombreuses analyses réalisées ces dernières années ont révélé des pressions de sélection variées ayant agi sur différents gènes de l'immunité innée (pour une revue des travaux, se référer à (Quintana-Murci and Clark, 2013). Par exemple, notre laboratoire a montré dès 2009 que certains récepteurs de type Toll (TLR), localisés dans les endosomes, étaient particulièrement conservés entre l'Homme et le Chimpanzé, suggérant une importance capitale de ces protéines dans la réponse mise en place suite à l'infection (Barreiro et al., 2009). À l'inverse, les TLR localisés à la membrane cytoplasmique sont plus

permissifs à l'accumulation de mutations non synonymes, suggérant qu'ils ont des fonctions redondantes ou qu'ils ont participé à l'adaptation de l'Homme. L'évolution des gènes impliqués dans la réponse immunitaire innée semble donc complexe, résultant de l'action de différentes pressions de sélection avec une intensité dépendant de la fonction biologique des protéines.

L'ensemble des études ayant pour but de disséquer l'histoire évolutive de ces gènes chez l'Homme ont été basées sur des gènes ou familles de gènes candidats. L'étude la plus large n'a été menée que sur les séquences codant pour des molécules impliquées spécifiquement dans la réponse antibactérienne. Une approche pangénomique de la manière dont la sélection et l'adaptation ont participé à l'évolution des gènes de l'immunité n'a jamais été réalisée. De plus, des données récentes indiquent que la diversité génétique observée dans l'espèce humaine résulte, en partie, d'un flux génétique d'hommes archaïques. Aucune analyse n'a, à ce jour, étudié l'influence que cette acquisition d'allèles Néandertaliens a eu sur l'évolution des gènes de l'immunité innée chez l'Homme.

Notre étude avait donc pour but de mettre à profit les données de séquençage du projet 1000 Génomes afin d'estimer dans quelle mesure la sélection naturelle avait agi sur l'ensemble des gènes impliqués dans l'immunité innée. Dans un premier temps, nous avons souhaité identifier des protéines pour lesquelles toute variation est délétère pour l'organisme. Dans un second temps, nous avons cherché à mettre en évidence des gènes ayant participé à l'adaptation de certaines populations humaines à leur environnement. Enfin, nous avons déterminé le rôle qu'a eu le flux de gènes Néanderthaliens dans l'évolution des gènes impliqués dans l'immunité innée.

## 1.2. Article

# Genomic Signatures of Selective Pressures and Introgression from Archaic Hominins at Human Innate Immunity Genes

Matthieu Deschamps,[1,2,3] Guillaume Laval,[1,2] Maud Fagny,[1,2,3] Yuval Itan,[4] Laurent Abel,[4,5,6] Jean-Laurent Casanova,[4,5,6,7,8] Etienne Patin,[1,2] and Lluis Quintana-Murci[1,2,*]

[1]Institut Pasteur, Unit of Human Evolutionary Genetics, 75015 Paris, France

[2]CNRS URA3012, 75015 Paris, France

[3]Université Pierre et Marie Curie, Cellule Pasteur UPMC, 75015 Paris, France

[4]St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, The Rockefeller University, New York, NY 10065, USA

[5]Laboratory of Human Genetics of Infectious Diseases, Necker Branch, INSERM U.1163, 7515 Paris, France

[6]Paris Descartes University, Imagine Institute, 75015 Paris, France

[7]Howard Hughes Medical Institute, New York, NY 10065, USA

[8]Pediatric Hematology-Immunology Unit, Necker Hospital for Sick Children, 75015 Paris, France

*Correspondence: quintana@pasteur.fr

## Abstract

Human genes governing innate immunity provide a valuable tool for the study of the selective pressure imposed by microorganisms on host genomes. A comprehensive, genome-wide study of how selective constraints and adaptations have driven the evolution of innate immunity genes is missing. Using full-genome sequence variation from the 1000 Genomes Project, we first show that innate immunity genes have globally evolved under stronger purifying selection than the remainder of protein-coding genes. We identify a gene set under the strongest selective constraints, mutations in which are likely to predispose individuals to life-threatening disease, as illustrated by *STAT1* and *TRAF3*. We then evaluate the occurrence of local adaptation, and detect 57 high-scoring signals of positive selection at innate immunity genes, variation in which has been associated with susceptibility to common infectious or autoimmune diseases. Furthermore, we show that most adaptations targeting coding variation have occurred in the last 6,000-13,000 years, the period at which populations shifted from hunting and gathering to farming. Finally, we show that innate immunity genes present higher Neanderthal introgression than the remainder of the coding genome. Notably, among the genes presenting the highest Neanderthal ancestry, we find the *TLR6-1-10* cluster, which also contains functional adaptive variation in Europeans. This study identifies highly constrained genes that fulfill essential, non-redundant functions in host survival while reveals others that are more permissive to change — containing variation acquired from archaic hominins or adaptive variants in specific populations — improving our understanding of the relative biological importance of innate immunity pathways in natural conditions.

# Introduction

The burden of infectious diseases has been massive throughout human history, particularly before the advent of hygiene, vaccines, antiseptics, and antibiotics, when human populations were ravaged by illnesses that resulted in high childhood mortality and short life expectancy.[1,2] In light of this, and given that the human genetic makeup strongly influences an individual's susceptibility to infectious disease and the resulting clinical outcome,[3-5] natural selection imposed by pathogens is expected to have profoundly affected the patterns of variability of the human genome.[6-10] Indeed, interspecies analyses and within-species studies in humans have established that purifying and positive selection have been both pervasive among genes and functions related to immunity and host defense at the genome-wide level.[7,11-17] Furthermore, pathogen pressure is increasingly recognized as the underlying cause of such selection signatures, with many immunity-related genes presenting patterns of variation that strongly correlate with pathogen diversity.[18-23]

Over recent decades, the dissection of the form and intensity of selection in the human genome has established the value of population genetics as a complement to clinical and epidemiological genetic studies, in delineating the biological relevance of immunity genes *in natura*, and in predicting their involvement in disease.[4,6,9,10,24-26] Genes evolving under strong purifying selection are predicted to be involved in essential mechanisms of host defense, variation in which should lead to severe disorders.[26] This prediction is supported by genome-wide data, as Mendelian disease genes are enriched in signals of purifying selection.[11,12,27] Conversely, genes evolving adaptively — through positive or balancing selection (e.g. *HBB*, *DARC*, *FUT2*, the *HLA* locus, ABO blood group, *TRIM5* genes) — are usually more permissive to functional variation, which can exert a protective effect against infections.[4,6,10,28] These signals of adaptive evolution in immune-related genes tending to be recent and population-specific further emphasizes the important role of pathogens in local adaptation.

Besides the occurrence of novel mutations, functional variants transmitted through admixture can represent another potential source of adaptive variation. Recent data provided evidence that 1–6% of modern Eurasian genomes were inherited from ancient hominins, such as Neanderthal or Denisovans,[29-31] with specific genomic regions presenting up to 64% of Neanderthal ancestry.[32] Furthermore, in the context of immunity, there is increasing evidence to suggest that modern humans have acquired advantageous variation through admixture with ancient hominins,[33] as documented by candidate gene approaches for *HLA* class I genes, *STAT2* or the *OAS* gene cluster.[34-36]

Among the two arms that constitute the immune system, innate immunity provides a valuable model for the study of the selective pressure imposed by microorganisms — pathogenic and symbiotic — on host genomes.[4,37] Innate immunity constitutes the front line of host defense, and relies on receptors that sense conserved microbial patterns or molecules.[38-40] Following ligand binding, receptors activate signaling pathways that involve the coordinated action of a diverse array of downstream molecules, including adaptors, regulators, transcription factors, and effector molecules, all of which are required for the eradication of pathogens and to maintain homeostasis.[40] Importantly, unlike adaptive immunity whose parameter variation is mostly somatic and presents limited heritability, variation of innate immunity is germline encoded and thus needs to be best-adapted to ensure host survival.[41,42] Population genetic studies have shown that the impact of selection on some families of innate immune receptors and downstream signaling molecules (e.g., Toll-like receptors, interferons, or antimicrobial peptides) varies considerably,[17,22,43-57] helping to delineate the relative functional importance of different immune pathways.[4,37] However, these studies have focused on specific candidate genes or gene families. A comprehensive, genome-wide view of how selection has driven the evolution of innate immunity in humans is thus missing.

Here, we took advantage of population whole-genome sequence data to increase our understanding of the degree of essentiality and adaptability of the different genes governing innate immunity and thus, to provide novel insights into their respective biological relevance in host survival. To do so, we first created a hand-curated list of more than 1,500 genes belonging to the different modules constituting the innate immune system in humans (for a detailed explanation of the definition and classification of innate immunity, see Material and Methods). We then analyzed their patterns of population genetic variability, which we compared to the remainder of the genome, using the 1000 Genomes Project dataset,[58] allowing us to evaluate the occurrence and intensity of constraint and adaptation to geographic and environmental pressures with an unprecedented level of resolution. Finally, we estimated the time range at which the bulk of genetic adaptation involving innate immunity has occurred as well as evaluated the extent to which human populations have acquired innate immunity genetic variation through admixture with Neanderthals.

# Material and Methods

## Hand-Curated List of Innate Immunity Genes

We created a hand-curated list of innate immunity genes (IIGs) by combining two public databases, Gene Ontology (GO)[59] and InnateDB[60], as well as by incorporating missing entries. Specifically, we used the GO term 'innate immune response" (GO:0045087) — defined as defense responses mediated by germline encoded components that directly recognize components of potential pathogens — to extract a list of 1,309 entries corresponding to 884 unique annotations (last access January 2015). We removed all non-human taxon entries, non-SwissProt reviewed proteins, entries without gene symbol or not approved by the HUGO Gene Nomenclature Committee, as well as those encoding for HLA proteins and immunoglobulins. This yielded a final set of 806 GO genes. As to InnateDB, it aims to facilitate systems level investigations of the innate immune response by curating human and mouse molecules, experimentally verified interactions and pathways involved in innate immunity, along with centralized annotation on the human and mouse interactomes.[61] For InnateDB, we retrieved a total of 2,158 entries, corresponding to 989 unique annotations (last access January 2015). Similarly to GO, we removed entries without approved HUGO names, *HLA* genes and miRNAs, and obtained a final set of 905 InnateDB genes.

When manually reviewing these two gene lists, we remarked (i) the absence of some proteins belonging, based on structural homology, to gene families involved in innate immunity, (ii) the absence of several well-described, or recently identified, molecules involved in innate immunity, and (iii) that some of these molecules were listed in databases with non-approved HUGO symbols. In light of this, for example, we added 28 TRIM proteins and 24 C-type lectins, some nucleic acid sensors such as *ABCF1*[62], *DHX15*[63], *DHX33*[64] or *PYHIN1*[65] as well as the interferons *IFNL1*, *IFNL2* and *IFNL3*, which were annotated in InnateDB as *IL29*, *IL28A* and *IL28B*, respectively. We acknowledge that some of the

molecules we have manually added, therefore absent from the curated lists that were downloaded from the public databases at the time of the study, have now been included in the corresponding websites. Overall, we manually added a set of 187 genes, making a final dataset of 1,553 genes that constituted the basis of all subsequent analyses (Table S1).

**Whole Genome Sequence Datasets**

For the purpose of this study, and depending on the nature of the analyses performed, we used the high-coverage (~57×) exome sequencing data as well as the low-coverage sequencing data (2-6×) of the 1000 Genomes Project, which are available for 1,092 individuals from 14 populations from Europe, East Asia, sub-Saharan Africa and the Americas.[58,66]

**Assessing the Action of Purifying Selection**

*Quantification of the Extent of Purifying Selection*

To estimate the strength of purifying selection, we used an approach, SnIPRE,[67] that relies on the comparison of polymorphism and divergence at synonymous and non-synonymous sites (i.e. McDonald-Kreitman contingency table). This method uses a generalized linear mixed model to model the genome-wide variability among categories of mutations and estimates two population genetics parameters for each gene: $\gamma$, the population selection coefficient, and $f$, the proportion of non-synonymous deleterious mutations. We focused our analyses on $f$, which quantifies the strength of purifying selection, i.e. the proportion of deleterious alleles that have been removed from the general population. SniPRE requires six different statistics that should be retrieved for all genes of the tested species (humans): the number of divergent non-synonymous and synonymous mutations between the tested species and an outgroup species (chimpanzees), the number of polymorphic non-synonymous and synonymous mutations in the tested species, and the number of non-synonymous and synonymous sites in the coding

sequence (i.e., the number of bases within a gene that generate a non-synonymous or synonymous mutation, if mutated).

To obtain these statistics in all human genes, we retrieved the alignment of the human genome (hg19 release) and the chimpanzee genome (PanTro3 release) provided by the UCSC Genome Browser, corresponding to ~2.5 Gb of aligned sequences. All regions of the human genome that are deleted or have no homology with the chimpanzee were excluded from the SniPRE analysis. We identified 33.5 million single bases that were different between the two species, which were then functionally annotated with snpEff,[68] using the GRCh37.65 build. We obtained 200,676 non-synonymous or synonymous divergent differences between humans and chimpanzees. We next retrieved all human variants that have been identified by the 1000 Genomes Project high-coverage exome dataset. We kept 445,401 variants that were annotated as non-synonymous or synonymous, were outside of gaps in the human-chimpanzee alignment and were polymorphic in at least one human population. Variants with a fixed alternate allele in the 1000 Genomes Project dataset (i.e., reference allele is absent from the sample) were added to fixed differences between human and chimpanzee. We excluded from human-chimpanzee fixed differences 16,345 positions that were actually polymorphic in humans or in chimpanzees, using the dbSNP136 chimpanzee database. Finally, we retrieved all human CDS with length >68bp and considered only the longest transcript available for each gene. We deduced from the genetic code the number of synonymous and non-synonymous sites in the 22,571 transcripts obtained, accounting for gaps in the human-chimpanzee alignments. All transcripts that had a length < 50bp after accounting for these gaps or had no divergent nor polymorphic mutations were excluded. SnIPRE[67] was then used to estimate the $f$ parameter for 18,997 genes, which included a final set of 1,492 IIGs, assuming human and chimpanzee sample sizes of 1,092 and 10, respectively.

*Statistical Analyses*

We tested for enrichments of IIGs among genes evolving under purifying selection by measuring the odds ratios (OR) of purifying selection. This OR measures the relative proportion of IIGs among genes with purifying selection signals and is defined as follows:

$$OR = \left[\frac{P(IIG|SEL)}{P(\overline{IIG}|SEL)}\right]\left[\frac{P(\overline{IIG}|\overline{SEL})}{P(IIG|\overline{SEL})}\right],$$

with $IIG$ and $\overline{IIG}$ denoting genes being or not innate immunity genes, respectively, $SEL$ and $\overline{SEL}$ being "with" and "without purifying selection signals", respectively. If purifying selection preferentially targets IIGs, we expect proportionally more IIGs in the tail of the *f* distribution (OR>1). Otherwise (i.e., purifying selection targets IIGs in the same manner as all other genes), we expect proportionally the same amount of IIGs in the tail of the *f* distribution as in the remainder of the genome (OR~1). Note that all statistical tests comparing *f* distributions among classes of genes (e.g., IIGs against the rest of human genes) were performed using resampling and classical tests assuming statistical independence between genes. Indeed, we observed a very weak correlation between *f* values of neighboring genes ($R^2$=0.016384), illustrating a limited hitchhiking effect in the context of purifying selection.

*Prediction of the Functional Impact of Mutations*

To evaluate the fitness status of variants at IIGs, we used the Combined Annotation Dependent Depletion (CADD) algorithm.[69] We downloaded the PHRED-scaled C-score calculation for the 39,701,210 variants (SNPs and InDels) from the 1000 Genomes Project and filtered out mutations that were excluded from the analyses of purifying selection. We then compared the number of SNPs in IIGs (33,907) and in the remainder of protein-coding sequences (411,307) having a PHRED-scaled score ≥ 15. We considered this value as the limit above which mutations are probably damaging, as this score corresponds to the median value for all possible canonical splice site changes and non-synonymous variants.[69]

*Protein-Protein Interaction Network Analysis*

We reconstructed the Protein-Protein interaction network by retrieving the interactions from the BioGRID database version 3.2.105 (ref.[70]). We retrieved protein Ensembl IDs using BioMart and considered only non-redundant direct physical interactions to compute degree centrality using the NetworkAnalyzer plugin[71] in Cytoscape.[72] Ubiquitin C and Amyloid Precursor Protein were removed from further analysis as they display outlier degree centralities. We transformed the degree centrality to $\log_{10}$(1+degree centrality) to reduce the skewness of the distribution, and used a Pearson correlation test to evaluate its relationship with the SnIPRE *f* parameter. We computed this correlation for the 1,114 IIGs and for the remaining 8,535 protein-coding sequences for which both degree centrality and *f* could be determined. We next compared these correlations using a linear model to estimate the effect of innate immunity in the relationship between *f* and degree centrality. For the representation of the innate immunity network, we used our IIG list as input for Cytoscape, and retrieved interactions among innate immunity proteins only using the MiMI plugin.[73]

*Transcription, Signal Transduction and Innate Immunity*

We retrieved the list of genes coding for proteins involved in transcription from Gene Ontology "transcription, DNA-templated" entry (GO:0006351). From this list of 2,643 genes, we extracted the 2,386 genes for which *f* values were calculated using SnIPRE. We considered genes at the intersection of this GO list and our set of IIGs as involved in both innate immune response and transcription. We then compared the distribution of *f* values between this group of genes involved in both innate immunity and transcription with that of genes involved only in transcription processes. Because our set of IIGs also includes entries from InnateDB, we performed the same analyses by restricting the comparisons only between

the two Gene Ontology terms “transcription, DNA-templated” and “innate immune response”. The same rationale was applied to the comparisons involving signal transducers. Specifically, we retrieved from Gene Ontology “intracellular signal transduction” entry (GO:0035556) and from the list of 1,875 unique entries, we extracted the 1,713 genes for which *f* values were available with SnIPRE.

**Genome-Wide Detection of Positive Selection**

*Detection of Positive Selection Using a Composite Statistics*

We combined, for each SNP, the set of statistics used in previous studies[7,74]: three based on haplotype homozygosity (iHS[16], ΔiHH[74] and XP-EHH[15]), and two based on the degree of population differentiation (ΔDAF[74] and $F_{ST}$[75]). In addition, we incorporated the DIND statistics,[17] which has been recently found to be powerful to detect positive selection using low-coverage whole-genome sequencing data.[76] For statistics based on haplotype homozygosity, we used the phased data of each population of the 1000 Genomes Project, and sliding windows of 100 kb centered on each SNP. This procedure does not alter the power to detect selection, while ensuring each statistics to be computed using equivalent regions, in terms of recombination rate, coverage and allele frequency spectrum.[76] Because some of these statistics require the ancestral/derived state of mutations, we retained sliding windows for which the ancestral/derived state of the core SNPs was unambiguously determined, i.e. 97% of the mutations of the 1000 Genomes dataset. Finally, we aimed to minimize the false positive rate, by excluding windows in which the core SNP had a derived allele frequency (DAF) below 0.2, as the power to detect selection at this allele frequency is limited.[76]

All these neutrality statistics were then combined into a Fisher's combined score ($F_{CS}$):

$$F_{CS} = -2\sum_{i=1}^{K} \ln(p_i)$$

where $K$ is the number of combined statistics and $p_i$ the empirical $P$-value for the $i^{th}$ statistics, i.e., the genomic rank of this $i^{th}$ statistics divided by the total number of unique values obtained for this statistics in the entire genome (values exactly equal get the same rank and same $P$-value). When $p_i$ values tend to be small, the $F_{CS}$ tends to be large. Under neutrality, $F_{CS}$ has a chi-squared distribution with $2K$ degrees of freedom. However, since the assumption of dependency among $p_i$ is violated, we used the genomic distribution of $F_{CS}$ in an empirical genome-wide test of selection, where the candidate SNPs with signals of selection are the ones exhibiting the 1% highest $F_{CS}$ values, as previously reported for other statistics.[16,76] Note that the $F_{CS}$ is computed for each population separately.

*Statistical Analyses*

Enrichments in positive selection signals among specific SNP classes (e.g., genic, located in IIGs, etc.) were tested as previously described.[11,16,76] Specifically, we used a logistic regression, generating an odds ratio (OR) for the effect of recent positive selection. For a given SNP class, the OR is defined as follows:

$$OR = \left[\frac{P(class|SEL)}{P(\overline{class}|SEL)}\right]\left[\frac{P(\overline{class}|\overline{SEL})}{P(class|\overline{SEL})}\right],$$

with $class$ being the SNP class, e. g., $class$ and $\overline{class}$ being genic and non genic SNPs respectively, $SEL$ and $\overline{SEL}$ being "with" and "without positive selection signal" respectively, i.e. SNP with an extreme $F_{CS}$ value. For example, if positive selection has preferentially occurred in genic regions, an OR>1 would be expected, reflecting the enrichment of the class of genic SNPs among SNPs with extreme $F_{CS}$ values (e.g., OR=1.25 when there are 20% true and 80% false positive among genic SNPs with extreme $F_{CS}$ values, see[76]). Otherwise (i.e., 100% of false positives among genic SNP outliers), we would expect an OR~1, indicating that the proportion of genic SNPs among outliers is not greater than the expected proportion of genic SNPs among all SNPs (~38% for the 1000 Genomes Project datasets, see[76]). The $P$-

values of enrichment analysis were obtained from 10,000 independent resamplings, taking into account linkage disequilibrium between SNPs.[76] For each resampling, we drew non-overlapping regions of 500 consecutive SNPs and arbitrarily assigned them to a given class, until we reached the number of SNPs observed in this SNP class. We considered the remaining SNPs to be out of the given class and calculated the OR for each resampling.[76]

*Identification of Candidate Regions*

To identify candidate genes, or gene regions, evolving adaptively, we used a conservative approach based on the degree of clustering of SNPs with extreme $F_{CS}$ values (i.e., the 1% top $F_{CS}$ values).[16,76] We used sliding windows of 100 kb centered on each SNP that contain at least 100 variants. We computed, for each 100 kb window, the proportion of extreme $F_{CS}$ values and grouped these windows into 75 bins of equal sizes based on the total number of SNPs observed. Finally, we considered the 1% of windows with the highest proportion of extreme $F_{CS}$ values in each bin as being under positive selection. A gene is thus considered to be a target of positive selection if it contains at least one window falling into this criterion.

*Assessing the Power to Detect Selection*

We evaluated the power to detect positive selection using the $F_{CS}$ statistics with computer simulations. We used cosi2 (ref.[77]) to simulate DNA regions according to realistic, accepted scenarios of human demography, as previously used for the 1000 Genomes Project dataset (for specific details on the parameters of the demographic model used, see ref.[7]). We simulated 60 unrelated individuals in each population sample, matching the 1000 Genomes Project dataset. We simulated 200-kb DNA regions with recombination rates sampled from the HapMap recombination map to generate realistic recombination patterns including local hotspots.[78] We simulated neutrally-evolving regions (with no selected site included in the

200-kb regions), and positive selection assuming the hard sweep model.[79] Specifically, a single new advantageous mutation with frequency 1/2N was inserted into the middle of the sequence in a specific population (YRI, CEU or CHB) at a specific time *t,* with a population genetics selection parameter 2N*s*=100 (selection coefficient *s*=0.01, N=10,000). We simulated different models of hard sweeps, by specifying various ages *t* of the selected allele (5 kya, 10 kya, 20 kya and 30 kya) and various $p_{sel}$, i.e. the frequency of the selected allele in the current generation (0.2, 0.4, 0.6, 0.8, and 1.0). We simulated 1,000 neutral-evolving regions and 100 regions for each combination of selection parameters (*t*, *s* and $p_{sel}$).

Since we used 100 kb windows centered on each SNP in the real data, and to avoid any truncation of these windows, we trimmed all simulated SNPs located at less than 50 kb of the edges of the 200 kb simulated regions. We computed the $F_{CS}$ statistics for each retained SNP located in the 100 kb in the middle of the 200 kb simulated region. We normalized iHS and DIND using the same method as previously described.[16,76] For each statistics, the empirical *P*-value $p_i$ used in the computation of $F_{CS}$ (see equation above) was determined for each population separately, by using all neutral simulations. We detected simulated regions under positive selection on the basis of the proportion of extreme $F_{CS}$ values. The power to detect positive selection was then computed as the proportion of regions simulated under positive selection effectively detected by our statistics (i.e., the percentage of simulations presenting proportions of extreme $F_{CS}$ values above the neutral threshold defined for a FPR of 1%).

*Annotation using hits of genome-wide association studies*

For each of the 57 IIGs presenting signatures of positive selection in our analysis (i.e. innate immunity genes carrying at least one SNP whose window has a proportion of outlier $F_{CS}$ among the 1% of genome-wide windows), we sought to explore their involvement in human diseases or traits using hits of genome-wide association studies (GWAS), obtained from the

02/06/2015 version of the NHGRI database. Only GWAS signals with *P*-values lower than $5x10^{-8}$ were considered. We used two approaches: the first gene-based approach relies on the simple fact that the tested IIG gene is the reported gene of a GWAS hit. The second SNP-based approach considers equivalence or strong linkage disequilibrium between SNPs candidate for positive selection and SNPs reported as best GWAS hits. For this second LD-aware approach, we selected all outlier SNPs (i.e., $F_{CS}$ among the 1% of genome-wide windows) in the genomic region of the tested IIG gene. We then retrieved all SNPs in strong linkage disequilibrium ($r^2$>0.8) with any of these candidate SNPs, using the correlation coefficient implemented in *plink*[80] on the unphased 1000 Genomes Project data for the relevant population (i.e. the population where the $F_{CS}$ signal was maximal). We finally checked whether some of our SNPs candidate for positive selection, or any SNP in LD with them, were among GWAS best signals.

**ABC Estimation of the Age of Selection**

To date the age of candidate mutations under positive selection, we used an approximate Bayesian computation (ABC) approach[81] to estimate the posterior probability of the age of selection according to the positive selection model described above: a single new advantageous mutation, occurring at a frequency of 1/2N, in a specific population, at a specific time *t* (i. e., the age of selection). We simulated 200-kb regions with a selected SNP located in the center of the sequence, according to the demographic model and recombination patterns described above. To generate a set of $2x10^5$ simulations, we used uniform prior distributions for the age and the intensity of selection and for the current frequency of the selected allele: the age of selection ($t$) varies from 0 to 62,500 years, the intensity of selection ($s$) varies from 0.002 to 0.05 and the current frequency of the selected allele ($p_{sel}$) from 0.2 to 1.0. Note that the prior distributions of the age and intensity of selection do not remain

uniform because some parameters vectors ($t$, $s$, $p_{sel}$) are unlikely (e.g., ancient selective events of strong intensity cannot generate a frequency of the selected allele equal to 0.2). Finally, we also simulated the low-coverage nature of the data (5×) by randomly drawing limited numbers of reads, as previously described.[76]

We started by using a set of summary statistics ($\Theta_S$, $\Theta_\pi$, Tajima's $D$, Fay and Wu's $H$, iHS, and $F_{ST}$) that has been previously found to be informative for estimating age of selection in an ABC framework,[82] to which we also incorporated the DIND statistics.[17,76] As performed for iHS and DIND, $\Theta_S$, $\Theta_\pi$, Tajima's $D$, and Fay and Wu's $H$ were computed in a window of 100 kb around the selected mutation. We used and tested the performance of various sets of summary statistics with different ABC methods — 'ridge' and 'neuralnet'— implemented in the 'abc' R package. We validated the performance of the ABC methods by using simulated datasets as if they were true empirical data, for which parameter values to estimate are known. This procedure allowed us to compare the estimated to the true values using various classical accuracy indices ('abc' R package): the prediction error *PE* (i.e. the mean square error, MSE, divided by the prior variance of the parameter), the relative estimation bias *rEB* (i.e. the bias expressed a proportion of the true value, also known as relative error) and the coverage of the 95% credible interval 95%COV (i.e. the percent of times where the true value was found within the 95% credible interval). Let $\theta_k$ and $\widehat{\theta_k}$ be the true and the ABC estimated values of the parameter $\theta$ in the $k^{th}$ simulated dataset:

$$PE = \frac{\frac{1}{S}\sum_1^S (\widehat{\theta_k} - \theta_k)^2}{var(\theta)},$$

$$rEB = \frac{1}{S}\sum_1^S \left(\widehat{\theta_k} - \theta_k\right)/\theta_k,$$

$$95\%COV = \frac{1}{S}\sum_1^S 1(q_1 < \theta_k < q_2\ ),$$

where $S$ is the number of simulated data, $1(C)$ the indicative function (equal to 1 when $C$ is true, 0 otherwise) and $q_1$ and $q_2$ the two percentiles of the posterior distribution of $\widehat{\theta_k}$ ($q_1$ and

$q_2$ were adjusted to obtain $95\%COV$ approximately equal to 0.95). These accuracy indices were computed using *S* equal to 300 simulated data. Note that the posterior distributions were obtained by retaining the 1000 "best" simulations in the ABC procedure.

We next aimed to improve the ABC estimations by adding more summary statistics. We first used the summary statistics described above (referred to as "set 1 of summary statistics") as well as their corresponding average and proportion of 1% top values computed over 100 kb around each candidate variant (referred to as "set 2 of summary statistics"). Furthermore, we aimed to boost the ABC estimations by including arithmetic transformations of the used summary statistics.[83] Specifically, we applied to the set 2 of summary statistics the following transformation, $T\left(S_i S_{j\geq i}\right)$ with $S_i$ and $S_j$ the $i^{\text{th}}$ and $j^{\text{th}}$ summary statistics. This simple procedure generates a new set of summary statistics (referred to as "set 3 of summary statistics"). The use of the set 3 of summary statistics in the 'neuralnet' ABC method will be referred to as the 'boosted-neuralnet' ABC method.

**Analysis of Neanderthal Ancestry**

To investigate putative introgression from archaic hominins to modern humans at innate immunity genes, we used the probabilities of Neanderthal ancestry calculated for each SNP of the 1000 Genomes Project dataset.[32] Sankararaman and colleagues estimated these probabilities using a conditional random field method, which takes into account the allelic state at a SNP in non-African, Neanderthal and Yoruba individuals, the relative sequence divergence between these individuals and the consistency of haplotype lengths with estimated time of interbreeding with archaic humans.[32] The high-coverage sequencing data of the Altai Neanderthal genome was used as representing the archaic genome,[84] and the Yoruba genomes were used as references, as they are assumed to harbor no Neanderthal ancestry. We downloaded the inferred Neanderthal ancestry at each allele in the 1000 Genomes European

and East Asian populations. We used the reported combined results across the CEU, GBR, FIN, IBS and TSI populations as representing the European population (referred here as EUR) and CHB, CHS and JPT as representing the East Asian population (referred here as ASN).

We calculated the average introgression score for each protein-coding gene (i.e. removing Open Reading Frames and genes encoding for putative proteins) as the average of the marginal probabilities of Neanderthal ancestry for all bases of the gene. We then compared the distributions of the average introgression scores for our set of IIGs and the remainder of protein-coding genes. *P*-values between distributions were obtained from $10^6$ independent resamplings taking into account the genomic correlation of average introgression scores. To this end, we retrieved genomic regions showing high probability to be introgressed from Neanderthal for each population, defined as runs of SNPs which have a probability of Neanderthal ancestry >0.9. We calculated the median length of contigs (called $ml_c$) obtained by constructing a tiling path across confidently inferred Neanderthal haplotypes in each population as described in ref.[32]. We divided the genome in adjacent windows of length $ml_c$ and assessed the distribution of the number of protein-coding genes and IIGs by window. Resamplings were then carried out taking into account the distribution of number of IIGs by window. We used the resampled datasets to obtain the expected distribution of Mann-Whitney *U* under the null hypothesis and calculate the empirical *P*-values.

Finally, we determined the 5% of genes harboring the highest probability of Neanderthal ancestry at the genome-wide level and searched for those that were involved in innate immunity processes. These latter analyses were restricted to the CEU and CHB populations, as they were those used to detect positive selection in modern humans. For the analyses at the haplotype level, we retrieved confidently inferred Neanderthal haplotypes, i.e., runs of SNPs that present a probability of Neanderthal ancestry >0.9 (see ref.[32]).

# Results

## Building of the Innate Immunity Gene List

We established a list of genes involved in innate immunity by combining two publicly available databases, Gene Ontology (GO) and InnateDB,[59,60] which we manually curated according to a number of criteria listed in Material and Methods. This yielded a curated set of 806 genes from GO and 905 from InnateDB, 345 of which were overlapping between the two datasets (Figure 1A). Furthermore, we incorporated an additional set of 187 genes, which were missing from these datasets at the time of the study, leaving a final list of 1,553 innate immunity genes, IIGs (Table S1, and Material and Methods for details on the definition of "innate immunity"). We then classified all genes according to their main known (or inferred) function into nine different categories, ranging from sensors of microorganisms or danger signals to adaptor and effector molecules, and also included regulators of the signaling pathways and accessory molecules necessary for an efficient immune response (Figure 1B).

## Pervasive Signatures of Purifying Selection at Innate Immunity Genes

To define the degree of selective constraint at IIGs, we first investigated the extent to which purifying selection has acted on the different categories of IIGs since the divergence of human and chimpanzee lineages. To do so, we used the exome dataset from the 1000 Genomes Project,[58] and merged all individuals into a single group to focus on the human lineage as a whole. For all protein-coding sequences, we estimated the *f* parameter using SnIPRE,[67] which estimates the degree of selective constraints at each gene using polymorphism and divergence data at non-synonymous and synonymous sites (Material and Methods). The lower the *f* value, the stronger the deficit of non-synonymous mutations compared to synonymous variants, highlighting strong evolutionary constraints (Table S1).

We found that the distribution of the *f* parameter for IIGs was significantly skewed towards lower values ($P=7\times10^{-4}$, based on $10^5$ resamplings taking into account gene length and number of SNPs per gene). This indicates that, taken as a whole, genes involved in innate immune processes eliminate proportionally more non-synonymous variants than the remainder of protein-coding genes (Figure 2A). When restricting our analyses to genes presenting the lowest *f* values at the genome-wide level, we observed a systematic, significant enrichment in IIGs using different percentiles (Figure 2B). For example, when focusing on genes displaying the 1% lowest *f* values genome-wide, we observed a strong, significant enrichment in IIGs (OR=2.45, $P=3.47\times10^{-5}$), corresponding to the set of IIGs that have evolved under the strongest degree of purifying selection (Table 1). To test the functional impact of naturally-occurring variation at IIGs on protein structure or function, we evaluated the deleteriousness of exonic variants using the PHRED-scaled C-scores provided by CADD.[69] Interestingly, the proportion of variants with a scaled C-score ≥15 (i.e. among the ~3% most deleterious mutations of the genome) was lower in IIGs compared to non-IIGs (0.424 and 0.458, respectively; $P=8.1\times10^{-27}$, Figure S1), providing further support to the notion that IIGs have evolved under stronger evolutionary constraints.

We next assessed if the global signal of strong purifying selection detected at IIGs differed among genes with distinct functional roles in innate immunity. A strongly significant difference was detected (Kruskal-Wallis rank sum test $P=2.2\times10^{-16}$, Figure 2C). Molecules involved in signal transduction and transcription were those presenting the strongest selective constraints. Such strong constraints could attest the additional involvement of these genes in functions other than innate immunity. To test this hypothesis, we compared the *f* values of the "signal transduction" and "transcription" groups of IIGs with those of signal transducers and transcription factors that are not part of innate immunity processes (Materials and Methods). Innate immunity molecules involved in these processes presented significantly lower *f* values

than their respective comparison groups (Figure S2), suggesting that their involvement in innate immunity has further participated to the strong constraints acting on these genes. We also observed that sensor and effector molecules presented the greatest range of *f* values (Figure 2C), indicating that the degree of constraint affecting these categories varies considerably among their members. For example, when comparing the *f* values among the different sub-families of receptors, we found that the family of Cytosolic Nucleic Acid Sensors (CNASs) displays the strongest deficit of non-synonymous mutations (Kruskal-Wallis rank sum test $P=4.6\times10^{-3}$), while RIG-I-like receptors (RLRs) were those evolving under the most relaxed selective constraints (Figure S3).

Finally, we tested if the varying degree of selective constraints detected at the different IIGs could be partly explained by their localization in the protein-protein interaction network (PIN). We thus reconstructed the innate immunity PIN using the protein interactions from the BioGRID database.[70] As previously detected in other protein interaction networks,[56,85-87] we observed a negative correlation between degree centrality and *f* values, with genes located in the center of the network (mostly signal transducers and transcription factors) presenting the strongest selective constraints (Figure 2D). Notably, we observed stronger negative correlations for IIG products than for the remainder of protein-coding genes (R=-0.337 and -0.184, respectively; $P<0.05$), suggesting a crucial role of network topology in driving the evolution of genes involved in innate immunity.

**Identification of Regions Presenting High-Confidence Signals of Positive Selection**

We next searched for IIGs that present signals of recent, population-specific positive selection, as they should contain variation that has contributed to human adaptation to varying environments. To do so, we restricted our analyses to (i) one population per geographic region — Yoruba from Nigeria (YRI), Northern Europeans (CEU), Han Chinese from Beijing

(CHB) — and (ii) the low-coverage dataset of the 1000 Genomes Project, as we needed to go beyond exonic regions to compute statistics that use the degree of haplotype homozygosity over large physical distances (Materials and Methods). Note that the low-coverage (~5×) of this dataset has been shown to have little impact on the power of some statistics to detect positive selection.[76]

To detect robust signals of local adaptation, we used a composite method, the Fisher's combined score ($F_{CS}$), as combining multiple tests at each SNP has been shown to increase power to detect positive selection, while minimizing the detection of false positive signals.[7,74] We combined statistics based on extended haplotype homozygosity (iHS, ΔiHH and XP-EHH), intra-allelic haplotype diversity (DIND) and population differentiation (ΔDAF and $F_{ST}$). We first assessed the power of the $F_{CS}$ by conducting a simulation-based study (Material and Methods, Figure S4), and found its power to be comparable to that of the composite of multiple signals (CMS) test.[7] We also found that the power of $F_{CS}$ declines with the age of selection, indicating that this score favors the detection of recent positive selection.

At the genome-wide level, among SNPs displaying multiple selection signals (i.e., those exhibiting extreme $F_{CS}$ values), we found a significant enrichment in genic with respect to non-genic regions (OR>1, $P<10^{-4}$, Table S2), as previously reported.[7,76] That no significant enrichment in SNPs located in IIGs was observed among SNPs with multiple selection signals, nor among the different categories of innate immunity genes (data not shown), suggests that positive selection has not targeted IIGs to a greater extent than the remainder of the genome (Table S2). However, we identified a set of IIGs presenting strong, robust signatures of positive selection, by looking at gene regions with a high clustering of SNPs presenting selection signals.[7,16,76] Specifically, we searched for 100-kb windows with the highest (top 1%) proportions of SNPs with extreme $F_{CS}$ values (Material and Methods), and found 1,110, 670 and 1,229 of such sliding windows, corresponding to 21, 16, and 22 genes,

in YRI, CEU, and CHB populations, respectively (Table 2). Notably, we retrieved several already reported signals of positive selection, including the *TLR6-1-10* gene cluster, *IL4*, *IFIH1*, *CD36* or *CEACAM1*,[7,11,17,22,88-91] but also a number of novel hits (Table 2).

To fine-map the candidate variants underlying the positive selection signals, we merged the SNPs from the genome-wide low-coverage and exome high-coverage datasets for each of the 57 candidate genes, considering also their flanking regions (1 Mb upstream and downstream of the corresponding genes). Indeed, the incorporation of the exome data allows detecting variants that have failed to pass the quality control filters and be missing in the low-coverage data (e.g., the non-synonymous SNP in *TLR1,* rs5743618, known to have a functional effect on NK-κB activity[17]). We then re-computed the $F_{CS}$ statistics on this merged dataset and determined the variants that exhibited the strongest selection signals (1% variants with highest $F_{CS}$; dark blue dots in Figure 3). Focusing on coding variation, we identified 13 high-scoring variants (12 non-synonymous variants and one stop mutation) in 11 genes (Table 2 and Figure 3), some of which have been previously identified as adaptive mutations, e.g., rs5743618 in *TLR1*[17], rs10930046 in *IFIH1*[22,88] or rs3211938 in *CD36*[90,91].

Finally, we sought to explore the involvement of our 57 candidate genes in human diseases or traits using GWAS databases. We found that 27 of them — for both known and newly-identified targets of selection — have been associated, to different extents, with common diseases, including susceptibility to infections or autoimmune disorders (enrichment resampling $P$=3.77x10$^{-4}$, 3.71x10$^{-2}$ and 0.058 in YRI, CEU, and CHB populations, respectively, compared to all IIGs; Table S3). For 13 of these genes, we also identified a strong correlation between our candidate SNPs for positive selection and GWAS best hits (Table S4). We therefore provide a list of high-confidence genes and mutations, many of which are novel candidates of selection and can be related to various phenotypic traits, which may have conferred a selective advantage for local adaptation to specific human populations.

**Estimating the Age of Genetic Adaptations Targeting Innate Immunity**

We next aimed to estimate the age, *t*, at which positive selection has targeted the high-scoring coding variants described above (Table 2), using an Approximate Bayesian Computation (ABC) framework.[81] We first checked the accuracy of the ABC estimations and tested the performance of various sets of summary statistics (Materials and Methods and ref.[82]) with different ABC methods (i.e., 'ridge' and 'neuralnet'), using simulated datasets (Table S5, Materials and Methods). As expected in such ABC settings,[82] we observed some overestimations of *t*, e. g., the best relative estimation bias (rEB) being around 0.3 when using the "set 2 of summary statistics" (Table S5). We thus aimed to improve our ABC estimations and found that the best estimations were obtained when boosting the ABC estimations by including arithmetic transformations of the summary statistics used.[83] Specifically, when implementing such a 'boosted-neuralnet' method, we obtained the greatest accuracy, i.e., lowest relative estimation bias (rEB) and lowest prediction error (PE) (see "set 3 of summary statistics" in Table S5). We also noticed that the estimations of *t* are more accurate for more recent events of positive selection, as expected when considering that the power to detect selection decreases with the age of selection (Figure S5 and Table S6).

When using the 'boosted-neuralnet' method to estimate the age of selection for the C/T-13910 polymorphism (rs4988235) in the *LCT* region, the most iconic case of positive selection in Europeans, which is associated with lactase-persistence in adulthood,[92-94] we found an estimated age of selection (7,100 years; 95%CI: 3,500-11,000, Table S7 and Figure S6) in good agreement with previous estimates (8,000 years in ref.[94], 7,400 years in ref.[93] and 11,200 in ref.[82]). Given such accurate estimates, we applied this procedure to estimate the ages of selection for the 13 high-scoring coding variants detected in our set of IIGs. In all cases, we dated ages of selection events at ~6,000-13,000 years ago (Table S7 and Figures

S7-S9) with a few exceptions. The most recent events of selection were estimated at less than 3,900 years, for *CD36* (a member of the scavenger receptor type B family) in Africans and *NRG1* (a membrane glycoprotein that mediates cell-cell signaling) in Asians, while we found that the age of the selected allele at *CLEC3B* (a member of the C-type lectin receptor family) in Asians was of 35,500 years (Table S7).

**Investigating Neanderthal Ancestry of Innate Immunity Genes**

Recent studies of individual loci have shown that immunity-related genes, whether adaptive immunity genes such as *HLA*, or innate immunity genes such as *STAT2* and *OAS*, carry distinct haplotypes in modern humans that appear to have introgressed from archaic populations, likely conferring, in some cases, a selective advantage to modern humans.[34-36] In light of this, we aimed to evaluate, at the genome-wide level, the extent to which modern humans have acquired variation at innate immunity genes via admixture with archaic humans. To do so, we first assessed the degree of Neanderthal ancestry among IIGs as a whole, taking advantage of the recently published Neanderthal introgression map.[32] We found that IIGs have a higher average introgression score when compared to the remainder of the coding genome, in both European and Asian groups ($P$=8×10$^{-6}$ and $P$=2×10$^{-6}$, respectively; Figure 4A and Material and Methods). Notably, these results were also significant when considering the different European and Asian subpopulations individually ($P$≤1.8×10$^{-5}$ in all subpopulations). Importantly, this result cannot be accounted for by the strong selective constraint detected at IIGs, as this selective regime has been associated to a decrease in Neanderthal ancestry.[32]

Next, we determined the 5% of genes harboring the highest probability of Neanderthal ancestry at the genome-wide level in each population, and searched for those that were involved in innate immunity. Out of the sets of 810 genes presenting the highest introgression

scores, we found 76 and 78 IIGs in European and Asian populations, respectively, 28 of which being shared between the two groups (Table S8). Among these genes, we found the *OAS* gene cluster in Europeans and Asians, as previously established in a candidate gene study.[36] Importantly, we detected additional genomic regions involved in innate immunity, including genes encoding receptors such as *NLRC5* in Asians, transcription factors such as *IRF6* in Asians, or effector molecules such as the *IFITM1-3* gene family in Europeans and some type-I IFNs in Asians (Table S8).

Remarkably, two genomic regions that we found to have high Neanderthal ancestry, the *TLR6-1-10* gene cluster and *SIRT1*, were also part of our high-confidence genes evolving under positive selection in European and Asian populations, respectively (Table 2). For these genomic regions, we identified the SNPs that were most probably introgressed from Neanderthal,[32] (Figures 4B and 4C) and determined if they were correlated to candidate SNPs for positive selection. In *SIRT1*, introgressed haplotypes were not specifically carrying any of our candidate SNPs, suggesting that variation acquired by archaic admixture at this locus is not adaptive. For the *TLR6-1-10* gene cluster, haplotypes of inferred Neanderthal origin (16% in CEU and 49% in CHB) were tagged by the SNPs detected as targets of positive selection in Europeans, including the non-synonymous *TLR1* SNP rs4833095, which is in partial linkage disequilibrium ($r^2$=0.657) with the functional rs5743618 variant. However, the rs4833095 allele most probably introgressed from Neanderthal is not the putatively selected derived allele (associated with protection against asthma, allergy and hay fever)[95-97] but the ancestral, rare allele. These patterns suggest a much more complex history than a single adaptive mutation transmitted to Eurasians on a Neanderthal haplotype background (Figures 4B and 4C).

## Discussion

In this study, we have taken advantage of whole genome sequence datasets to provide a comprehensive assessment of how selection, in its different forms and intensities, has driven the evolution of innate immunity genes in humans. We must first point out that any definition, including ours, of "innate immunity genes" is arbitrary and not exempt of ambiguity. There is a clear distinction between innate and adaptive immunity at the cellular level but the distinction is blurred at the genetic level because many genes govern both innate and adaptive immunity.[98] Herein, our definition of innate immunity includes intrinsic, non-hematopoietic immunity, in addition to the traditional definition of innate immunity as hematopoietic and non-adaptive. Furthermore, for the choice of innate immunity genes, we merged two well-established public databases, which we manually curated and completed with missing hits. Bearing this in mind, there are several, important insights that can be drawn from our study.

First, we show that innate immunity genes, taken as a whole, evolve under stronger selective constraints than the remainder of protein-coding genes, indicating that the purge of deleterious mutations has been particularly important in this gene class. This observation is consistent with pathogens being one of the most important long-time threats to human survival throughout evolution.[6,23] Furthermore, innate immunity is germline-encoded, unlike adaptive immunity whose variation is mostly somatic,[41,42] and ensures the sensing of pathogens, inducing tightly controlled responses upon activation, and the maintenance of homeostasis with symbiotic microbiota.[38-40] Consequently, any mutation disturbing these processes would be deleterious and rapidly eliminated from the population. However, the strength of selective constraints varies considerably among different functional categories, as well as among the different members within each, informing us about the degree of redundancy or essentiality of the corresponding genes. For example, among microbial sensors, the cytosolic RLRs evolve under the most relaxed selective constraints, attesting to

higher immunological redundancy,[51] whereas the group of CNAS evolves under the strongest purifying selection, similarly to the group of endosomal TLRs,[17] suggesting that variation at these molecules might be strongly deleterious for the host.

That genes evolving under strong purifying selection are likely to fulfill essential, non-redundant functions in host defense, and be involved in severe clinical phenotypes, is well illustrated by the cases of *STAT1* and *TRAF3*. These two genes are among the 1% most constrained of the human genome (Table 1), and mutations in these genes have been shown to underlie severe primary immunodeficiencies.[4] Indeed, gain-of-function and loss-of-function mutations at STAT1, a protein involved in the transduction of cellular responses to IFN-α/β, -λ and γ, and IL-27,[99] have been associated with a range of immunological and clinical phenotypes, including life-threatening and mild bacterial (mainly mycobacterial) and viral (mainly herpes) diseases, Mendelian susceptibility to mycobacterial disease, chronic mucocutaneous candidiasis and autoimmunity.[100] Similarly, deficiency in TRAF3, which functions downstream of the TNF receptors, and IFN-inducing receptors including endosomal TLRs and RLRs,[101] has been associated with an immunological phenotype — the impairment of TLR3-dependent induction of IFN — that leads to herpes simplex virus 1 encephalitis, a devastating infection of the central nervous system.[102] These examples support the notion that the genes we report as targeted by strong purifying selection are of major biological relevance in host survival. Importantly, because of the pleiotropic functions ensured by many innate immunity genes and the broad definition we used to define them, some of these genes can be involved in mechanisms that can go beyond immunity to infection, including housekeeping functions. Regardless of the breadth of their biological functions, mutations in highly constrained IIGs are likely to predispose individuals to life-threatening disease; combining next-generation sequencing and evolutionary data in clinical studies should facilitate the discovery of novel genetic etiologies of severe, infectious disease phenotypes.

Second, our study represents the first attempt to characterize genome-wide signatures of positive selection at innate immunity genes using a composite metrics that improves power and resolution. We show that innate immunity genes have not undergone hard sweeps to a greater extent than the remainder of the genome, supporting the notion that polygenic adaptation — subtle adaptive effects at multiple genes — has been pervasive among functions related to innate immunity.[103] However, our analyses identify 57 regions presenting high-confidence signals of classic sweeps in specific human populations (Table 2). Interestingly, our age estimations show that most adaptations targeting coding variation at innate immunity genes have occurred in the last 6,000-13,000 years. It is noteworthy that given the nature of the selective signals we detect, we expect some over-representation of "young" selection events (e.g., recombination erases the hitchhiking effect with time). However, it is remarkable that the estimated age ranges correspond very well with the transition from food collection (hunting/gathering) to food production (farming/herding), which occurred in many parts of the world starting 10-13,000 years ago.[104] The shift to agriculture led humans to adopt sedentary lifestyles and was accompanied by increased population density, food crises and contacts with cattle and biological wastes, which modified human exposure to pathogens[105] and, as our results suggest, was associated to some degree of genetic adaptation.

Several of our high-scoring positively-selected genes have been reported to be associated, through GWAS or candidate-gene studies, with common infectious or autoimmune phenotypes (Table S3), and contain variants that have been previously detected as evolving adaptively in specific populations; e.g., *IFIH1* (inflammatory bowel disease, vitiligo, type-1 diabetes, psoriasis) and *CD36* (malaria) in Africans, the *TLR6-1-10* cluster (allergy, asthma, leprosy) in Europeans, or *BLK* (rheumatoid arthritis, systemic lupus erythematosus, Kawasaki disease) in Asians.[7,11,17,22,88-91] The case of the stop mutation at *CD36* (T1264G; rs3211938) is particularly worth discussing. CD36 is an archetypal pattern recognition receptor that binds

polyanionic ligands of both pathogen and self origin, promotes inflammation in monocytes and macrophages through a TLR4-TLR6 heterodimer, and mediates cytoadherence of *Plasmodium falciparum* parasitized erythrocytes.[106-108] Although the association between *CD36* and severe malaria remains complex,[90,109-112] the stop variant represents a well-supported case of strong positive selection,[90,91,113,114] which our analysis clearly confirms, reaching a frequency of 29% in the Yoruba from Nigeria. It has been proposed that the high frequency of the 1264G allele in Nigeria, together with its low frequencies in the rest of Africa, results from a geographically confined selective event.[90] We estimate the age of the selected 1264G mutation at only 3,600 years (95% CI: 2,125-5,025 years), the youngest selective event detected by our analyses, strongly supporting the notion that the increase in frequency of this mutation restricted to West-Central Africa represents a local, recent and strong event of genetic adaptation. Investigating the association between *CD36* and severe malaria phenotypes specifically in Nigerian populations is now needed.

Importantly, our high-score list of positively selected genes includes novel hits that have not been previously investigated as targets of selection but contain SNPs that have been associated with immunity-related phenotypes (Table S3). Notably, we have detected two high-scoring non-synonymous mutations in *MERTK* (rs7604639 and rs2230515, Figure 3C), with a derived allele frequency of 79% in the Asian population. *MERTK* is a member of the 3 TAM receptor tyrosine kinases that are involved in the regulation of inflammatory responses,[115] and variation in this gene has been associated with hepatitis C-induced liver fibrosis.[116] Likewise, our analysis suggests that positive selection has increased the frequency of a non-synonymous mutation in *ZFPM2* (rs11993776) to 58% in Africans. ZFPM2 modulates the activity of GATA transcription factors at the *HAMP* promoter, an antimicrobial peptide involved in the metabolism of iron,[117] which is critical for *Mycobacterium tuberculosis* growth in macrophages and proposed to be a risk factor for developing

tuberculosis.[118]. Interestingly, variation at *ZFPM2* has been recently suggested to be associated with susceptibility to tuberculosis in a South African admixed population.[119] Altogether, we provide a tractable list of high-scoring selected coding variants for experimental follow-up, which are likely to have played a dominant role in recent adaptations of human populations to their respective environments.

Finally, our study provides new insight into the degree of putative introgression of innate immunity genes from archaic hominins. It has been shown that protein-coding genes are generally depleted in Neanderthal ancestry, owing to the widespread effects of negative selection against Neanderthal ancestry in gene regions.[32] Interestingly, we find that innate immunity genes present both stronger evidence for purifying selection and a higher average probability of Neanderthal ancestry than the remainder of the coding genome, indicating a weaker purge, or slightly stronger selective advantage, of Neanderthal alleles in innate immunity genes in Eurasian populations. For example, among the genes presenting the highest inferred Neanderthal ancestry we find the IFITM1-3 proteins (Table S8), a family of restriction factors that are highly inducible by IFN-α/β and IFN-γ and restrict the replication of multiple pathogenic viruses *in vitro*, including influenza A virus, dengue virus and West Nile virus.[120,121] In particular, variation at *IFITM3* has been suggested to alter the morbidity and mortality associated with influenza infection in humans.[122] The case of *IFITM3* illustrates more generally the higher tolerance observed in innate immunity genes for Neanderthal introgression, suggesting that archaic admixture may have constituted a source of advantageous variation for host defense in modern humans.

Although neutral introgression appears to be the most likely explanation for most genes, the case of the *TLR6-1-10* cluster is particularly worth discussing. First, it presents high Neanderthal introgression scores in both Europeans and Asians (Figure 4B and Table S8). Second, the *TLR6-1-10* cluster has been proposed to be a hotspot of positive selection, given

the strong signals of selective sweeps detected in human and non-human primates.[123] Third, we find this region, here and elsewhere, as being a strongly supported case of local adaptation in Europeans (Table 2 and refs.[17,55,124]). Furthermore, three high-scoring adaptive non-synonymous mutations have been detected in this gene cluster (Figure 3A), one of which (rs5743618 in *TLR1,* I602S) appears to be the genuine target of positive selection; it has been shown to remarkably impair agonist-induced NF-κB activation by up to 60% and be linked to infectious disease phenotypes, such as leprosy.[17,55,125-127] However, this adaptive hypo-responsiveness allele is not present in the Neanderthal genomes.[128] More generally, for SNPs most probably introgressed from Neanderthal that show signatures of positive selection in Europeans, the alleles observed in archaic hominins are rare and ancestral in modern humans, while positive selection has targeted the highly-frequent derived alleles. Altogether, while we provide compelling evidence supporting both high Neanderthal ancestry and positive selection for functional mutations at the *TLR6-1-10* cluster, our analyses show that Neanderthal introgression is probably not the source of such adaptive variation. Future studies should experimentally test whether the Neanderthal introgressed variation detected at this gene cluster has any functional impact on TLR-mediated immunity to infection.

In summary, our analyses have shown that the contemporary diversity of innate immunity genes in humans results from the intermingling of different demographic and selective events, including introgression from Neanderthal, hard sweeps at some loci in specific human populations occurring mostly during the Neolithic transition, and continued selective constraints at other loci. In doing so, they increase our understanding of the degree of essentiality and adaptability of innate immunity genes, with novel candidates for having played a dominant role in recent adaptations, and provide new insight into the extent to which modern humans may have acquired variation at innate immunity genes through admixture with archaic hominins.

## Supplemental Data

Supplemental data include nine figures and eight tables and can be found with this article online at http://www.cell.com/AJHG/.


## Acknowledgments

The authors thank Matthew Albert and Jean-Marc Cavaillon for useful advices and discussions. This work was supported by the Institut Pasteur, the *Centre Nationale de la Recherche Scientifique* (CNRS), and the *Agence Nationale de la Recherche* (ANR) grants: "DEMOCHIPS" ANR-12-BSV7-0012, "IEIHSEER " ANR-14-CE14-0008-02 and "TBPATHGEN" ANR-14-CE14-0007-02. The laboratory of LQM has received funding from the French Government's Investissement d'Avenir program, Laboratoire d'Excellence "Integrative Biology of Emerging Infectious Diseases" (grant no. ANR-10- LABX-62-IBEID), and from the European Research Council under the European Union's Seventh Framework Programme (FP/2007–2013)/ERC Grant Agreement No. 281297.


## Web Resources

The URLs for data presented herein are as follows:

Combined Annotation Dependent Depletion (CADD), http://cadd.gs.washington.edu/

Gene Ontology Consortium, http://geneontology.org/

The NHGRI/EMBL Catalog of Published Genome-Wide Association Studies, http://www.genome.gov/gwastudies/

InnateDB: Systems Biology of the Innate Immune Response, http://www.innatedb.com/

The 1000 Genomes Project, http://www.1000genomes.org/

The landscape of Neanderthal ancestry in present-day humans, http://genetics.med.harvard.edu/reichlab/Reich_Lab/Datasets_-_Neandertal_Introgression.html

Tools for Approximate Bayesian Computation (ABC), http://cran.r-project.org/web/packages/abc/index.html

# References


1. Cairns, J. (1997) Matters of Life and Death. Princeton University Press, Princeton, NJ
2. Casanova, J.L., Abel, L. (2005). Inborn errors of immunity to infection: the rule rather than the exception. J Exp Med *202*, 197-201.
3. Abel, L., Alcais, A., Schurr, E. (2014). The dissection of complex susceptibility to infectious disease: bacterial, viral and parasitic infections. Curr Opin Immunol *30*, 72-78.
4. Casanova, J.L., Abel, L., Quintana-Murci, L. (2013). Immunology taught by human genetics. Cold Spring Harb Symp Quant Biol *78*, 157-172.
5. Chapman, S.J., Hill, A.V. (2012). Human genetic susceptibility to infectious disease. Nat Rev Genet *13*, 175-188.
6. Barreiro, L.B., Quintana-Murci, L. (2010). From evolutionary genetics to human immunology: how selection shapes host defence genes. Nat Rev Genet *11*, 17-30.
7. Grossman, S.R., Andersen, K.G., Shlyakhter, I., Tabrizi, S., Winnicki, S., Yen, A., Park, D.J., Griesemer, D., Karlsson, E.K., Wong, S.H., et al. (2013). Identifying recent adaptations in large-scale genomic data. Cell *152*, 703-713.
8. Fumagalli, M., Sironi, M. (2014). Human genome variability, natural selection and infectious diseases. Curr Opin Immunol *30C*, 9-16.
9. Siddle, K.J., Quintana-Murci, L. (2014). The Red Queen's long race: human adaptation to pathogen pressure. Curr Opin Genet Dev *29C*, 31-38.
10. Karlsson, E.K., Kwiatkowski, D.P., Sabeti, P.C. (2014). Natural selection and infectious disease in human populations. Nat Rev Genet *15*, 379-393.
11. Barreiro, L.B., Laval, G., Quach, H., Patin, E., Quintana-Murci, L. (2008). Natural selection has driven population differentiation in modern humans. Nat Genet *40*, 340-345.

12. Bustamante, C.D., Fledel-Alon, A., Williamson, S., Nielsen, R., Hubisz, M.T., Glanowski, S., Tanenbaum, D.M., White, T.J., Sninsky, J.J., Hernandez, R.D., et al. (2005). Natural selection on protein-coding genes in the human genome. Nature *437*, 1153-1157.

13. Leffler, E.M., Gao, Z., Pfeifer, S., Segurel, L., Auton, A., Venn, O., Bowden, R., Bontrop, R., Wall, J.D., Sella, G., et al. (2013). Multiple instances of ancient balancing selection shared between humans and chimpanzees. Science *339*, 1578-1582.

14. Nielsen, R., Hellmann, I., Hubisz, M., Bustamante, C., Clark, A.G. (2007). Recent and ongoing selection in the human genome. Nat Rev Genet *8*, 857-868.

15. Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R., et al. (2007). Genome-wide detection and characterization of positive selection in human populations. Nature *449*, 913-918.

16. Voight, B.F., Kudaravalli, S., Wen, X., Pritchard, J.K. (2006). A map of recent positive selection in the human genome. PLoS Biol *4*, e72.

17. Barreiro, L.B., Ben-Ali, M., Quach, H., Laval, G., Patin, E., Pickrell, J.K., Bouchier, C., Tichit, M., Neyrolles, O., Gicquel, B., et al. (2009). Evolutionary dynamics of human Toll-like receptors and their different contributions to host defense. PLoS Genet *5*, e1000562.

18. Prugnolle, F., Manica, A., Charpentier, M., Guegan, J.F., Guernier, V., Balloux, F. (2005). Pathogen-driven selection and worldwide HLA class I diversity. Curr Biol *15*, 1022-1027.

19. Pozzoli, U., Fumagalli, M., Cagliani, R., Comi, G.P., Bresolin, N., Clerici, M., Sironi, M. (2010). The role of protozoa-driven selection in shaping human genetic variability. Trends Genet *26*, 95-99.

20. Fumagalli, M., Pozzoli, U., Cagliani, R., Comi, G.P., Bresolin, N., Clerici, M., Sironi, M. (2010). The landscape of human genes involved in the immune response to parasitic worms. BMC Evol Biol *10*, 264.

21. Fumagalli, M., Pozzoli, U., Cagliani, R., Comi, G.P., Bresolin, N., Clerici, M., Sironi, M. (2010). Genome-wide identification of susceptibility alleles for viral infections through a population genetics approach. PLoS Genet *6*, e1000849.

22. Fumagalli, M., Cagliani, R., Riva, S., Pozzoli, U., Biasin, M., Piacentini, L., Comi, G.P., Bresolin, N., Clerici, M., Sironi, M. (2010). Population genetics of IFIH1: ancient population structure, local selection, and implications for susceptibility to type 1 diabetes. Mol Biol Evol *27*, 2555-2566.

23. Fumagalli, M., Sironi, M., Pozzoli, U., Ferrer-Admetlla, A., Pattini, L., Nielsen, R. (2011). Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. PLoS Genet *7*, e1002355.

24. Manry, J., Laval, G., Patin, E., Fornarino, S., Itan, Y., Fumagalli, M., Sironi, M., Tichit, M., Bouchier, C., Casanova, J.L., et al. (2011). Evolutionary genetic dissection of human interferons. J Exp Med *208*, 2747-2759.

25. Quintana-Murci, L., Alcais, A., Abel, L., Casanova, J.L. (2007). Immunology in natura: clinical, epidemiological and evolutionary genetics of infectious diseases. Nat Immunol *8*, 1165-1171.

26. Alcais, A., Quintana-Murci, L., Thaler, D.S., Schurr, E., Abel, L., Casanova, J.L. (2010). Life-threatening infectious diseases of childhood: single-gene inborn errors of immunity? Ann N Y Acad Sci *1214*, 18-33.

27. Blekhman, R., Man, O., Herrmann, L., Boyko, A.R., Indap, A., Kosiol, C., Bustamante, C.D., Teshima, K.M., Przeworski, M. (2008). Natural selection on genes that underlie human disease susceptibility. Curr Biol *18*, 883-889.

28. Key, F.M., Teixeira, J.C., de Filippo, C., Andres, A.M. (2014). Advantageous diversity maintained by balancing selection in humans. Curr Opin Genet Dev *29C*, 45-51.

29. Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H., et al. (2010). A draft sequence of the Neandertal genome. Science *328*, 710-722.

30. Meyer, M., Kircher, M., Gansauge, M.T., Li, H., Racimo, F., Mallick, S., Schraiber, J.G., Jay, F., Prufer, K., de Filippo, C., et al. (2012). A high-coverage genome sequence from an archaic Denisovan individual. Science *338*, 222-226.

31. Reich, D., Green, R.E., Kircher, M., Krause, J., Patterson, N., Durand, E.Y., Viola, B., Briggs, A.W., Stenzel, U., Johnson, P.L., et al. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature *468*, 1053-1060.

32. Sankararaman, S., Mallick, S., Dannemann, M., Prufer, K., Kelso, J., Paabo, S., Patterson, N., Reich, D. (2014). The genomic landscape of Neanderthal ancestry in present-day humans. Nature *507*, 354-357.

33. Segurel, L., Quintana-Murci, L. (2014). Preserving immune diversity through ancient inheritance and admixture. Curr Opin Immunol *30*, 79-84.

34. Abi-Rached, L., Jobin, M.J., Kulkarni, S., McWhinnie, A., Dalva, K., Gragert, L., Babrzadeh, F., Gharizadeh, B., Luo, M., Plummer, F.A., et al. (2011). The shaping of modern human immune systems by multiregional admixture with archaic humans. Science *334*, 89-94.

35. Mendez, F.L., Watkins, J.C., Hammer, M.F. (2012). A haplotype at STAT2 Introgressed from neanderthals and serves as a candidate of positive selection in Papua New Guinea. Am J Hum Genet *91*, 265-274.

36. Mendez, F.L., Watkins, J.C., Hammer, M.F. (2013). Neandertal origin of genetic variation at the cluster of OAS immunity genes. Mol Biol Evol *30*, 798-801.

37. Quintana-Murci, L., Clark, A.G. (2013). Population genetic tools for dissecting innate immunity in humans. Nat Rev Immunol *13*, 280-293.

38. Beutler, B., Jiang, Z., Georgel, P., Crozat, K., Croker, B., Rutschmann, S., Du, X., Hoebe, K. (2006). Genetic analysis of host resistance: Toll-like receptor signaling and immunity at large. Annual review of immunology *24*, 353-389.

39. Medzhitov, R. (2007). Recognition of microorganisms and activation of the immune response. Nature *449*, 819-826.

40. Takeuchi, O., Akira, S. (2010). Pattern recognition receptors and inflammation. Cell *140*, 805-820.

41. Brodin, P., Jojic, V., Gao, T., Bhattacharya, S., Angel, C.J., Furman, D., Shen-Orr, S., Dekker, C.L., Swan, G.E., Butte, A.J., et al. (2015). Variation in the human immune system is largely driven by non-heritable influences. Cell *160*, 37-47.

42. Casanova, J.L., Abel, L. (2015). Disentangling inborn and acquired immunity in human twins. Cell *160*, 13-15.

43. Mukherjee, S., Sarkar-Roy, N., Wagener, D.K., Majumder, P.P. (2009). Signatures of natural selection are not uniform across genes of innate immune system, but purifying selection is the dominant signature. Proc Natl Acad Sci U S A *106*, 7073-7078.

44. Nakajima, T., Ohtani, H., Satta, Y., Uno, Y., Akari, H., Ishida, T., Kimura, A. (2008). Natural selection in the TLR-related genes in the course of primate evolution. Immunogenetics *60*, 727-735.

45. Wlasiuk, G., Khan, S., Switzer, W.M., Nachman, M.W. (2009). A history of recurrent positive selection at the toll-like receptor 5 in primates. Mol Biol Evol *26*, 937-949.

46. Wlasiuk, G., Nachman, M.W. (2010). Adaptation and constraint at Toll-like receptors in primates. Mol Biol Evol *27*, 2172-2186.

47. Ferrer-Admetlla, A., Bosch, E., Sikora, M., Marques-Bonet, T., Ramirez-Soriano, A., Muntasell, A., Navarro, A., Lazarus, R., Calafell, F., Bertranpetit, J., et al. (2008). Balancing selection is the main force shaping the evolution of innate immunity genes. J Immunol *181*, 1315-1322.

48. Ferrer-Admetlla, A., Sikora, M., Laayouni, H., Esteve, A., Roubinet, F., Blancher, A., Calafell, F., Bertranpetit, J., Casals, F. (2009). A natural history of FUT2 polymorphism in humans. Mol Biol Evol *26*, 1993-2003.

49. Fornarino, S., Laval, G., Barreiro, L.B., Manry, J., Vasseur, E., Quintana-Murci, L. (2011). Evolution of the TIR domain-containing adaptors in humans: swinging between constraint and adaptation. Mol Biol Evol *28*, 3087-3097.

50. Ferwerda, B., McCall, M.B., Alonso, S., Giamarellos-Bourboulis, E.J., Mouktaroudi, M., Izagirre, N., Syafruddin, D., Kibiki, G., Cristea, T., Hijmans, A., et al. (2007). TLR4 polymorphisms, infectious diseases, and evolutionary pressure during migration of modern humans. Proc Natl Acad Sci U S A *104*, 16645-16650.

51. Vasseur, E., Boniotto, M., Patin, E., Laval, G., Quach, H., Manry, J., Crouau-Roy, B., Quintana-Murci, L. (2012). The evolutionary landscape of cytosolic microbial sensors in humans. Am J Hum Genet *91*, 27-37.

52. Verdu, P., Barreiro, L.B., Patin, E., Gessain, A., Cassar, O., Kidd, J.R., Kidd, K.K., Behar, D.M., Froment, A., Heyer, E., et al. (2006). Evolutionary insights into the high worldwide prevalence of MBL2 deficiency alleles. Hum Mol Genet *15*, 2650-2658.

53. Ferwerda, B., Alonso, S., Banahan, K., McCall, M.B., Giamarellos-Bourboulis, E.J., Ramakers, B.P., Mouktaroudi, M., Fain, P.R., Izagirre, N., Syafruddin, D., et al. (2009). Functional and genetic evidence that the Mal/TIRAP allele variant 180L has

been selected by providing protection against septic shock. Proc Natl Acad Sci U S A *106*, 10272-10277.

54. Hollox, E.J., Armour, J.A. (2008). Directional and balancing selection in human beta-defensins. BMC Evol Biol *8*, 113.

55. Laayouni, H., Oosting, M., Luisi, P., Ioana, M., Alonso, S., Ricano-Ponce, I., Trynka, G., Zhernakova, A., Plantinga, T.S., Cheng, S.C., et al. (2014). Convergent evolution in European and Rroma populations reveals pressure exerted by plague on Toll-like receptors. Proc Natl Acad Sci U S A *111*, 2668-2673.

56. Casals, F., Sikora, M., Laayouni, H., Montanucci, L., Muntasell, A., Lazarus, R., Calafell, F., Awadalla, P., Netea, M.G., Bertranpetit, J. (2011). Genetic adaptation of the antibacterial human innate immunity network. BMC Evol Biol *11*, 202.

57. Cagliani, R., Forni, D., Biasin, M., Comabella, M., Guerini, F.R., Riva, S., Pozzoli, U., Agliardi, C., Caputo, D., Malhotra, S., et al. (2014). Ancient and recent selective pressures shaped genetic diversity at AIM2-like nucleic acid sensors. Genome Biol Evol *6*, 830-845.

58. Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56-65.

59. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet *25*, 25-29.

60. Breuer, K., Foroushani, A.K., Laird, M.R., Chen, C., Sribnaia, A., Lo, R., Winsor, G.L., Hancock, R.E., Brinkman, F.S., Lynn, D.J. (2013). InnateDB: systems biology of innate immunity and beyond--recent updates and continuing curation. Nucleic Acids Res *41*, D1228-1233.

61. Lynn, D.J., Winsor, G.L., Chan, C., Richard, N., Laird, M.R., Barsky, A., Gardy, J.L., Roche, F.M., Chan, T.H., Shah, N., et al. (2008). InnateDB: facilitating systems-level analyses of the mammalian innate immune response. Mol Syst Biol *4*, 218.

62. Lee, M.N., Roy, M., Ong, S.E., Mertins, P., Villani, A.C., Li, W., Dotiwala, F., Sen, J., Doench, J.G., Orzalli, M.H., et al. (2013). Identification of regulators of the innate immune response to cytosolic DNA and retroviral infection by an integrative approach. Nat Immunol *14*, 179-185.

63. Lu, H., Lu, N., Weng, L., Yuan, B., Liu, Y.J., Zhang, Z. (2014). DHX15 senses double-stranded RNA in myeloid dendritic cells. J Immunol *193*, 1364-1372.

64. Mitoma, H., Hanabuchi, S., Kim, T., Bao, M., Zhang, Z., Sugimoto, N., Liu, Y.J. (2013). The DHX33 RNA helicase senses cytosolic RNA and activates the NLRP3 inflammasome. Immunity *39*, 123-135.

65. Schattgen, S.A., Fitzgerald, K.A. (2011). The PYHIN protein family as mediators of host defenses. Immunological Reviews *243*, 109-118.

66. Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., McVean, G.A. (2010). A map of human genome variation from population-scale sequencing. Nature *467*, 1061-1073.

67. Eilertson, K.E., Booth, J.G., Bustamante, C.D. (2012). SnIPRE: selection inference using a Poisson random effects model. PLoS Comput Biol *8*, e1002806.

68. Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) *6*, 80-92.

69. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet *46*, 310-315.

70. Chatr-Aryamontri, A., Breitkreutz, B.J., Oughtred, R., Boucher, L., Heinicke, S., Chen, D., Stark, C., Breitkreutz, A., Kolas, N., O'Donnell, L., et al. (2015). The BioGRID interaction database: 2015 update. Nucleic Acids Res *43*, D470-478.

71. Assenov, Y., Ramirez, F., Schelhorn, S.E., Lengauer, T., Albrecht, M. (2008). Computing topological parameters of biological networks. Bioinformatics *24*, 282-284.

72. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res *13*, 2498-2504.

73. Gao, J., Ade, A.S., Tarcea, V.G., Weymouth, T.E., Mirel, B.R., Jagadish, H.V., States, D.J. (2009). Integrating and annotating the interactome using the MiMI plugin for cytoscape. Bioinformatics *25*, 137-138.

74. Grossman, S.R., Shlyakhter, I., Karlsson, E.K., Byrne, E.H., Morales, S., Frieden, G., Hostetter, E., Angelino, E., Garber, M., Zuk, O., et al. (2010). A composite of multiple signals distinguishes causal variants in regions of positive selection. Science *327*, 883-886.

75. Holsinger, K.E., Weir, B.S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting F(ST). Nat Rev Genet *10*, 639-650.

76. Fagny, M., Patin, E., Enard, D., Barreiro, L.B., Quintana-Murci, L., Laval, G. (2014). Exploring the occurrence of classic selective sweeps in humans using whole-genome sequencing data sets. Mol Biol Evol *31*, 1850-1868.

77. Shlyakhter, I., Sabeti, P.C., Schaffner, S.F. (2014). Cosi2: an efficient simulator of exact and approximate coalescent with selection. Bioinformatics *30*, 3427-3429.

78. Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., et al. (2007). A second generation human haplotype map of over 3.1 million SNPs. Nature *449*, 851-861.

79. Pritchard, J.K., Pickrell, J.K., Coop, G. (2010). The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. Curr Biol *20*, R208-215.

80. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet *81*, 559-575.

81. Beaumont, M.A., Zhang, W., Balding, D.J. (2002). Approximate Bayesian computation in population genetics. Genetics *162*, 2025-2035.

82. Peter, B.M., Huerta-Sanchez, E., Nielsen, R. (2012). Distinguishing between selective sweeps from standing variation and from a de novo mutation. PLoS Genet *8*, e1003011.

83. Aeschbacher, S., Beaumont, M.A., Futschik, A. (2012). A novel approach for choosing summary statistics in approximate Bayesian computation. Genetics *192*, 1027-1047.

84. Prufer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P.H., de Filippo, C., et al. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. Nature *505*, 43-49.

85. Hahn, M.W., Kern, A.D. (2005). Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. Mol Biol Evol *22*, 803-806.

86. Fraser, H.B., Hirsh, A.E., Steinmetz, L.M., Scharfe, C., Feldman, M.W. (2002). Evolutionary rate in the protein interaction network. Science *296*, 750-752.

87. Lemos, B., Bettencourt, B.R., Meiklejohn, C.D., Hartl, D.L. (2005). Evolution of proteins and gene expression levels are coupled in Drosophila and are independently associated with mRNA abundance, protein length, and number of protein-protein interactions. Mol Biol Evol *22*, 1345-1354.

88. Vasseur, E., Patin, E., Laval, G., Pajon, S., Fornarino, S., Crouau-Roy, B., Quintana-Murci, L. (2011). The selective footprints of viral pressures at the human RIG-I-like receptor family. Hum Mol Genet *20*, 4462-4474.

89. Colonna, V., Ayub, Q., Chen, Y., Pagani, L., Luisi, P., Pybus, M., Garrison, E., Xue, Y., Tyler-Smith, C., Abecasis, G.R., et al. (2014). Human genomic regions with exceptionally high levels of population differentiation identified from 911 whole-genome sequences. Genome Biol *15*, R88.

90. Fry, A.E., Ghansa, A., Small, K.S., Palma, A., Auburn, S., Diakite, M., Green, A., Campino, S., Teo, Y.Y., Clark, T.G., et al. (2009). Positive selection of a CD36 nonsense variant in sub-Saharan Africa, but no association with severe malaria phenotypes. Hum Mol Genet *18*, 2683-2692.

91. Sabeti, P.C., Schaffner, S.F., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O., Palma, A., Mikkelsen, T.S., Altshuler, D., Lander, E.S. (2006). Positive natural selection in the human lineage. Science *312*, 1614-1620.

92. Bersaglieri, T., Sabeti, P.C., Patterson, N., Vanderploeg, T., Schaffner, S.F., Drake, J.A., Rhodes, M., Reich, D.E., Hirschhorn, J.N. (2004). Genetic signatures of strong recent positive selection at the lactase gene. Am J Hum Genet *74*, 1111-1120.

93. Itan, Y., Powell, A., Beaumont, M.A., Burger, J., Thomas, M.G. (2009). The origins of lactase persistence in Europe. PLoS Comput Biol *5*, e1000491.

94. Tishkoff, S.A., Reed, F.A., Ranciaro, A., Voight, B.F., Babbitt, C.C., Silverman, J.S., Powell, K., Mortensen, H.M., Hirbo, J.B., Osman, M., et al. (2007). Convergent adaptation of human lactase persistence in Africa and Europe. Nat Genet *39*, 31-40.

95. Bonnelykke, K., Matheson, M.C., Pers, T.H., Granell, R., Strachan, D.P., Alves, A.C., Linneberg, A., Curtin, J.A., Warrington, N.M., Standl, M., et al. (2013). Meta-analysis of genome-wide association studies identifies ten loci influencing allergic sensitization. Nat Genet *45*, 902-906.

96. Ferreira, M.A., Matheson, M.C., Tang, C.S., Granell, R., Ang, W., Hui, J., Kiefer, A.K., Duffy, D.L., Baltic, S., Danoy, P., et al. (2014). Genome-wide association analysis identifies 11 risk variants associated with the asthma with hay fever phenotype. J Allergy Clin Immunol. *133*, 1564-1571.

97. Hinds, D.A., McMahon, G., Kiefer, A.K., Do, C.B., Eriksson, N., Evans, D.M., St Pourcain, B., Ring, S.M., Mountain, J.L., Francke, U., et al. (2013). A genome-wide association meta-analysis of self-reported allergy identifies shared and allergy-specific susceptibility loci. Nat Genet *45*, 907-911.

98. Flajnik, M.F., Du Pasquier, L. (2004). Evolution of innate and adaptive immunity: can we draw a line? Trends Immunol *25*, 640-644.

99. van Boxel-Dezaire, A.H., Rani, M.R., Stark, G.R. (2006). Complex modulation of cell type-specific signaling in response to type I interferons. Immunity *25*, 361-372.

100. Boisson-Dupuis, S., Kong, X.F., Okada, S., Cypowyj, S., Puel, A., Abel, L., Casanova, J.L. (2012). Inborn errors of human STAT1: allelic heterogeneity governs the diversity of immunological and infectious phenotypes. Curr Opin Immunol *24*, 364-378.

101. Hacker, H., Tseng, P.H., Karin, M. (2011). Expanding TRAF function: TRAF3 as a tri-faced immune regulator. Nat Rev Immunol *11*, 457-468.

102. Perez de Diego, R., Sancho-Shimizu, V., Lorenzo, L., Puel, A., Plancoulaine, S., Picard, C., Herman, M., Cardon, A., Durandy, A., Bustamante, J., et al. (2010). Human TRAF3 adaptor molecule deficiency leads to impaired Toll-like receptor 3 response and susceptibility to herpes simplex encephalitis. Immunity *33*, 400-411.

103. Daub, J.T., Hofer, T., Cutivet, E., Dupanloup, I., Quintana-Murci, L., Robinson-Rechavi, M., Excoffier, L. (2013). Evidence for polygenic adaptation to pathogens in the human genome. Mol Biol Evol *30*, 1544-1558.

104. Diamond, J., Bellwood, P. (2003). Farmers and their languages: the first expansions. Science *300*, 597-603.

105. Wolfe, N.D., Dunavan, C.P., Diamond, J. (2007). Origins of major human infectious diseases. Nature *447*, 279-283.

106. Newbold, C., Craig, A., Kyes, S., Rowe, A., Fernandez-Reyes, D., Fagan, T. (1999). Cytoadherence, pathogenesis and the infected red cell surface in Plasmodium falciparum. Int J Parasitol *29*, 927-937.

107. Hoebe, K., Georgel, P., Rutschmann, S., Du, X., Mudd, S., Crozat, K., Sovath, S., Shamel, L., Hartung, T., Zahringer, U., et al. (2005). CD36 is a sensor of diacylglycerides. Nature *433*, 523-527.

108. Stewart, C.R., Stuart, L.M., Wilkinson, K., van Gils, J.M., Deng, J., Halle, A., Rayner, K.J., Boyer, L., Zhong, R., Frazier, W.A., et al. (2010). CD36 ligands promote sterile inflammation through assembly of a Toll-like receptor 4 and 6 heterodimer. Nat Immunol *11*, 155-161.

109. Malaria Genomic Epidemiology Network (2014). Reappraisal of known malaria resistance loci in a large multicenter study. Nat Genet *46*, 1197-1204.

110. Aitman, T.J., Cooper, L.D., Norsworthy, P.J., Wahid, F.N., Gray, J.K., Curtis, B.R., McKeigue, P.M., Kwiatkowski, D., Greenwood, B.M., Snow, R.W., et al. (2000). Malaria susceptibility and CD36 mutation. Nature *405*, 1015-1016.

111. Omi, K., Ohashi, J., Patarapotikul, J., Hananantachai, H., Naka, I., Looareesuwan, S., Tokunaga, K. (2003). CD36 polymorphism is associated with protection from cerebral malaria. Am J Hum Genet *72*, 364-374.

112. Pain, A., Urban, B.C., Kai, O., Casals-Pascual, C., Shafi, J., Marsh, K., Roberts, D.J. (2001). A non-sense mutation in Cd36 gene is associated with protection from severe malaria. Lancet *357*, 1502-1503.

113. International HapMap Consortium (2005). A haplotype map of the human genome. Nature *437*, 1299-1320.

114. Bhatia, G., Patterson, N., Pasaniuc, B., Zaitlen, N., Genovese, G., Pollack, S., Mallick, S., Myers, S., Tandon, A., Spencer, C., et al. (2011). Genome-wide comparison of African-ancestry populations from CARe and other cohorts reveals signals of natural selection. Am J Hum Genet *89*, 368-381.

115. Lemke, G., Rothlin, C.V. (2008). Immunobiology of the TAM receptors. Nat Rev Immunol *8*, 327-336.

116. Patin, E., Kutalik, Z., Guergnon, J., Bibert, S., Nalpas, B., Jouanguy, E., Munteanu, M., Bousquet, L., Argiro, L., Halfon, P., et al. (2012). Genome-wide association study identifies variants associated with progression of liver fibrosis from HCV infection. Gastroenterology *143*, 1244-1252 e1241-1212.

117. Bagu, E.T., Santos, M.M. (2011). Friend of GATA suppresses the GATA-induced transcription of hepcidin in hepatocytes through a GATA-regulatory element in the HAMP promoter. J Mol Endocrinol *47*, 299-313.

118. Boelaert, J.R., Vandecasteele, S.J., Appelberg, R., Gordeuk, V.R. (2007). The effect of the host's iron status on tuberculosis. J Infect Dis *195*, 1745-1753.

119. Chimusa, E.R., Zaitlen, N., Daya, M., Moller, M., van Helden, P.D., Mulder, N.J., Price, A.L., Hoal, E.G. (2014). Genome-wide association study of ancestry-specific TB risk in the South African Coloured population. Hum Mol Genet *23*, 796-809.

120. Lewin, A.R., Reid, L.E., McMahon, M., Stark, G.R., Kerr, I.M. (1991). Molecular analysis of a human interferon-inducible gene family. Eur J Biochem *199*, 417-423.

121. Brass, A.L., Huang, I.C., Benita, Y., John, S.P., Krishnan, M.N., Feeley, E.M., Ryan, B.J., Weyer, J.L., van der Weyden, L., Fikrig, E., et al. (2009). The IFITM proteins mediate cellular resistance to influenza A H1N1 virus, West Nile virus, and dengue virus. Cell *139*, 1243-1254.

122. Everitt, A.R., Clare, S., Pertel, T., John, S.P., Wash, R.S., Smith, S.E., Chin, C.R., Feeley, E.M., Sims, J.S., Adams, D.J., et al. (2012). IFITM3 restricts the morbidity and mortality associated with influenza. Nature *484*, 519-523.

123. Enard, D., Depaulis, F., Roest Crollius, H. (2010). Human and non-human primate genomes share hotspots of positive selection. PLoS Genet *6*, e1000840.

124. Pickrell, J.K., Coop, G., Novembre, J., Kudaravalli, S., Li, J.Z., Absher, D., Srinivasan, B.S., Barsh, G.S., Myers, R.M., Feldman, M.W., et al. (2009). Signals of recent positive selection in a worldwide sample of human populations. Genome Res *19*, 826-837.

125. Hawn, T.R., Misch, E.A., Dunstan, S.J., Thwaites, G.E., Lan, N.T., Quy, H.T., Chau, T.T., Rodrigues, S., Nachman, A., Janer, M., et al. (2007). A common human TLR1 polymorphism regulates the innate immune response to lipopeptides. European journal of immunology *37*, 2280-2289.

126. Johnson, C.M., Lyle, E.A., Omueti, K.O., Stepensky, V.A., Yegin, O., Alpsoy, E., Hamann, L., Schumann, R.R., Tapping, R.I. (2007). Cutting edge: A common polymorphism impairs cell surface trafficking and functional responses of TLR1 but protects against leprosy. J Immunol *178*, 7520-7524.

127. Misch, E.A., Macdonald, M., Ranjit, C., Sapkota, B.R., Wells, R.D., Siddiqui, M.R., Kaplan, G., Hawn, T.R. (2008). Human TLR1 deficiency is associated with impaired mycobacterial signaling and protection from leprosy reversal reaction. PLoS Negl Trop Dis *2*, e231.

128. Castellano, S., Parra, G., Sanchez-Quinto, F.A., Racimo, F., Kuhlwilm, M., Kircher, M., Sawyer, S., Fu, Q., Heinze, A., Nickel, B., et al. (2014). Patterns of coding variation in the complete exomes of three Neandertals. Proc Natl Acad Sci U S A *111*, 6666-6671.

# Figure Legends

**Figure 1. Hand-curated list of innate immunity genes.**

(**A**) Venn diagram representing the three sets of innate immunity genes used in this study. In red, the final set of 806 genes associated with the GO term "innate immune response" (GO:0045087); in green, the final set of 905 genes from InnateDB; and in blue, the final set of 1,553 innate immunity genes, which includes the 187 manually-added genes (for details, see Material and Methods). (**B**) Schematic representation of the classification of innate immunity genes into different functional categories, together with the number of genes assigned to each category. Note that each gene was given a unique assignation to a particular functional class.

**Figure 2. Varying degrees of selective constraints targeting innate immunity genes.**

(**A**) Strength of purifying selection acting on innate immunity genes and the remainder of protein-coding genes, as measured by the *f* value. We tested the significance of the observed difference by means of $10^5$ resamplings taking into account gene length and number of SNPs per gene in the two tested gene sets (*** $P=7\times10^{-4}$). (**B**) Enrichment of innate immunity genes among the most constrained genes at the genome-wide level, as assessed by odds ratios (ORs). We calculated ORs for increasing percentiles of the *f* distribution, with a pace of 1%. The 95% confidence intervals of ORs were calculated using the Fisher's exact test. (**C**) Strength of purifying selection acting on the different functional categories of innate immunity genes, as measured by the *f* value (UC stands for unclassified). (**D**) Innate immunity protein interaction network. Only innate immunity proteins interacting with a molecular partner also involved in this cellular process are represented. Node sizes are negatively correlated to *f* values, i.e., large nodes represent low *f* values, indicating stronger action of purifying selection. Color codes are the same as those used in (**C**).

**Figure 3. Innate immunity genes presenting high-confidence signals of geographic adaptation**

Four examples of innate immunity genes presenting strong signals of positive selection, including (**A**) the *TLR6-1-10* gene cluster, (**B**) *IFIH1*, (**C**) *MERTK* and (**D**) *ZFPM2*. The black curves delineate the proportions of outlier SNPs (i.e., SNPs with the 1% highest $F_{CS}$ values of the genome), within 100-kb regions, at the genome-wide level, using the low-coverage 1000 Genomes Project dataset (see Material and Methods for details). Blue dots represent the $F_{CS}$ value of each SNP, calculated using the merged dataset (both high- and low-coverage) for the fine-mapping of putative adaptive mutations. Dark blue dots indicate SNPs with the 1% highest $F_{CS}$ values of the genome, within which non-synonymous variants are represented by red triangles. The remaining variants are plotted in light blue, where triangles represent non-synonymous mutations.

**Figure 4. Neanderthal ancestry of innate immunity genes.**

(**A**) Comparison of the average introgression scores of innate immunity genes (IIGs) with respect to the remainder of protein-coding genes (non-IIGs) in European (EUR) and East Asian (ASN) populations. ***: $P$ <0.001 (see Material and Methods) (**B,C**) Haplotypes of Neanderthal ancestry in (**B**) CEU individuals at the *TLR6-1-10* gene cluster and (**C**) CHB individuals at the *SIRT1* locus. Confidently inferred haplotypes of Neanderthal ancestry, defined as long runs of SNPs that present a probability of Neanderthal ancestry >0.9 (ref.[32]), are indicated in blue in each diploid individual from the 1000 Genomes Project. Red shadows highlight genomic regions containing innate immunity genes.

**Table 1. Innate immunity genes presenting the strongest signatures of purifying selection at the genome-wide level**

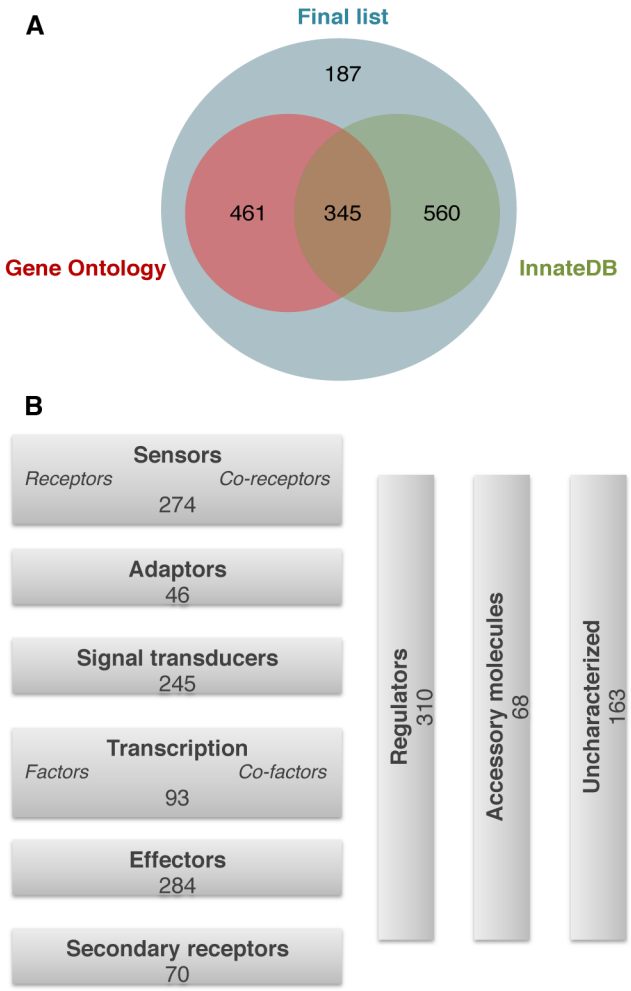
| Functional Category | Genes |
|---|---|
| Sensors | *DHX9* |
| Adaptors | *CNKSR2*, *CTNNB1*, *SRC*, *SYK* |
| Signal transducers | *CAMK2B*, *MTOR*, *TRAF3* |
| Transcription | *GATA3*, *HNRNPL*, *SMARCA2*, *SMARCA4*, *STAT1* |
| Effectors | *AGO1*, *AGO2*, *AGO3* |
| Secondary receptors | *EGFR* |
| Accessory molecules | *HSP90AB1*, *UBC* |
| Regulators | *CYLD*, *HDAC1*, *KHSRP*, *MID2*, *USP7* |
| UC | *ACTB*, *ACTG1*, *CYFIP2*, *DOCK1*, *FSCN1*, *ITPR1*, *NCKAP1*, *TUBB4B* |

**Table 2. Innate immunity genes showing the strongest signatures of positive selection**

| Population | Innate immunity genes[a] |
|---|---|
| YRI | *VSP45, CD1D, FCER1A, LTBP1, CCDC88A, LY75, IFIH1, LTF, CCR2, CD80, MAPK10, CD36*[b]*, ZFPM2, TRIM55, CHUK, DAK, POLR3B, HIF1A, CEACAM1, TNRC6B, MYH9* |
| CEU | *CCDC88A, TLR10, TLR1, TLR6, MAP3K1, IL4, IRGM, TRIM27, EYA4, ARPC1A, ZC3HAV1, SRPK2, SMARCA2, SIRT1, DUOX1, ADAM10* |
| CHB | *ARHGEF2, ADAM15, LYST, PELI1, ACTR2, MERTK, ERBB4, SP100, RAF1, LRRFIP2, CLEC3B, RHOA, GAB1, FER, ITPR3, EGFR, BLK, NRG1, SIRT1, OTUB1, ARHGEF7, PIAS4* |

[a]These genes overlap with at least one 100-kb window with the 1% highest proportions of outlier SNPs with the highest $F_{CS}$ values. Genes that are underlined contain at least one non-synonymous mutation with an outlier $F_{CS}$ value: *IFIH1*: rs10930046; *LTF*: rs60938611; *ZFPM2*: rs11993776; *DAK*: rs2260655; *TLR1*: rs5743618 and rs4833095; *TLR6*: rs5743810; *MAP3K1*: rs702689; *MERTK*: rs7604639 and rs2230515; *CLEC3B*: rs13963; *NRG1*: rs3924999. [b]*CD36* contains an outlier premature stop mutation (rs3211938).

1 Figure 1


A
Final list
187
461
345
560
Gene Ontology
InnateDB
B
Sensors
Receptors
Co-receptors
274
Adaptors
46
Signal transducers
245
Transcription
Factors
Co-factors
93
Effectors
284
Secondary receptors
70
Regulators
310
Accessory molecules
68
Uncharacterized
163

2

3

Figure 2

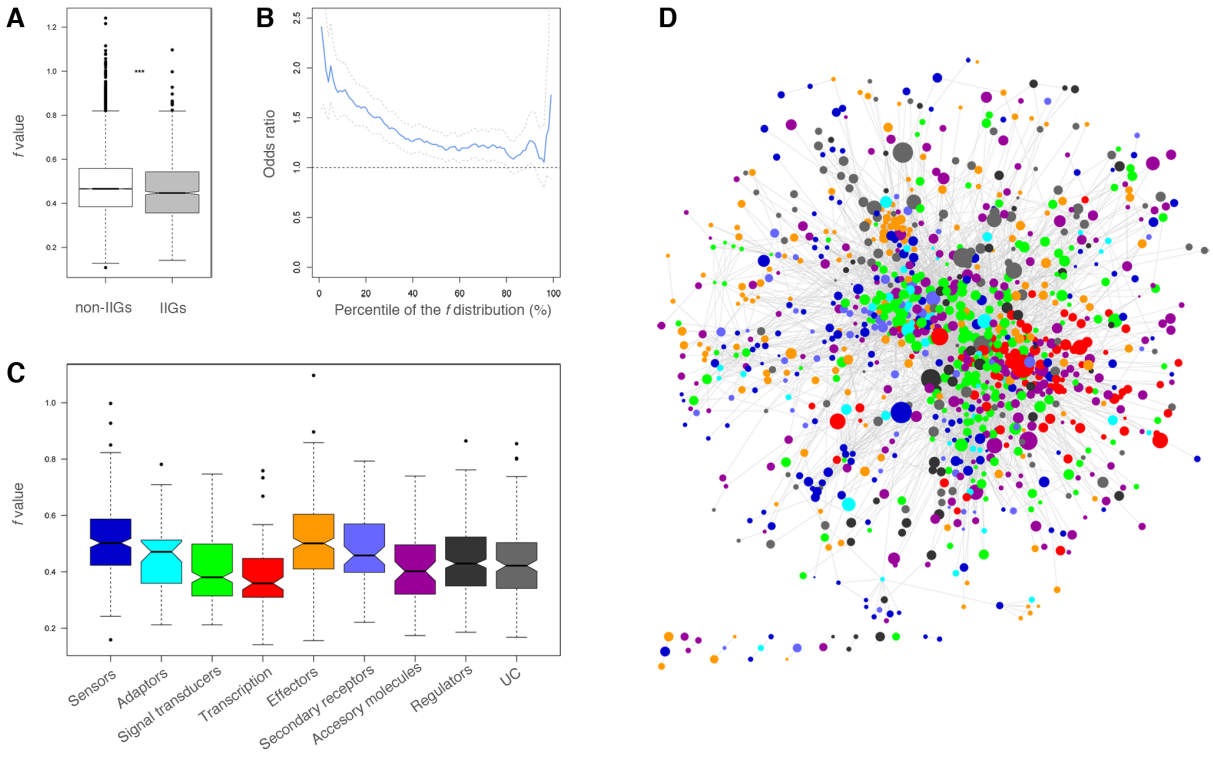
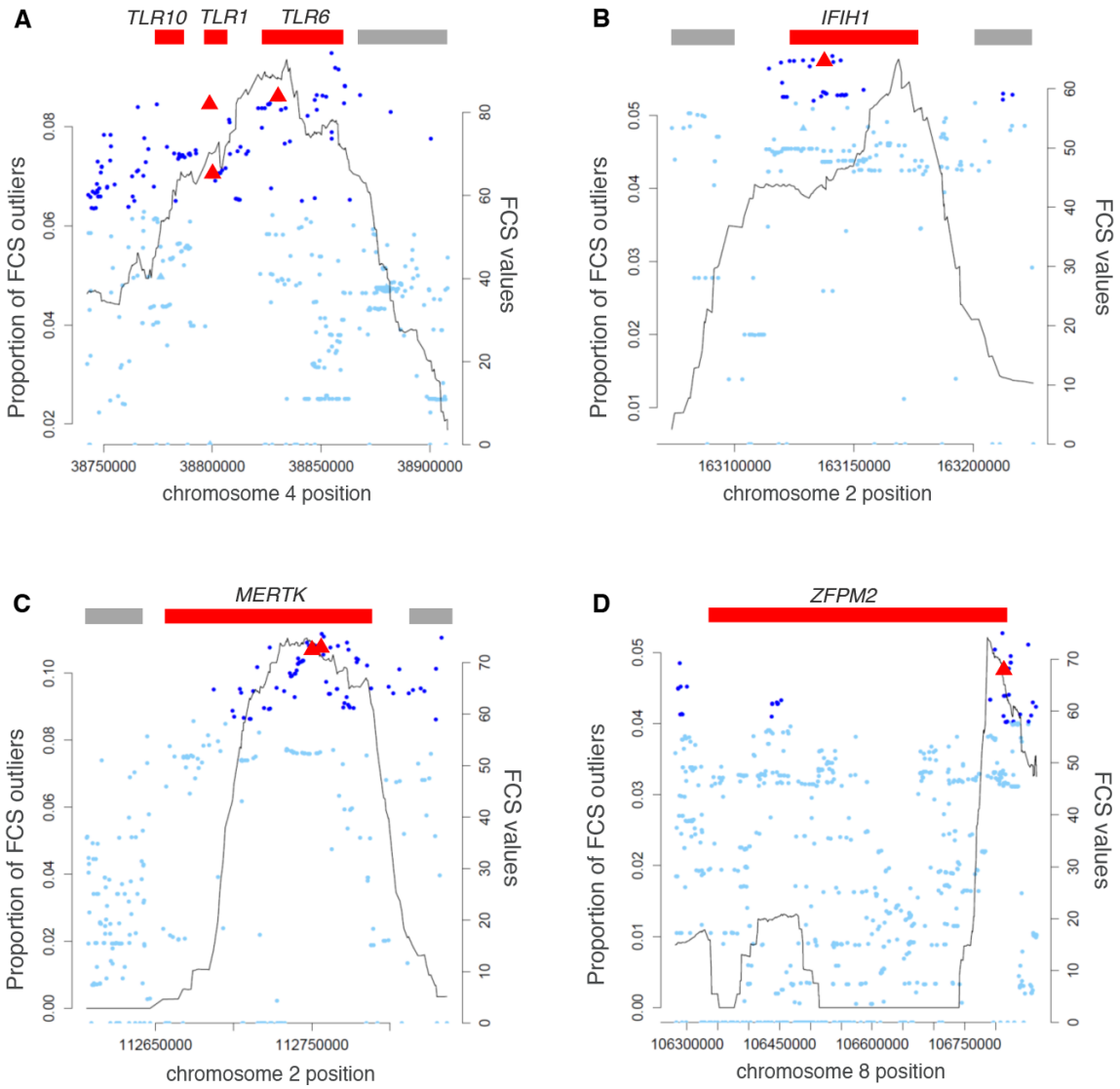
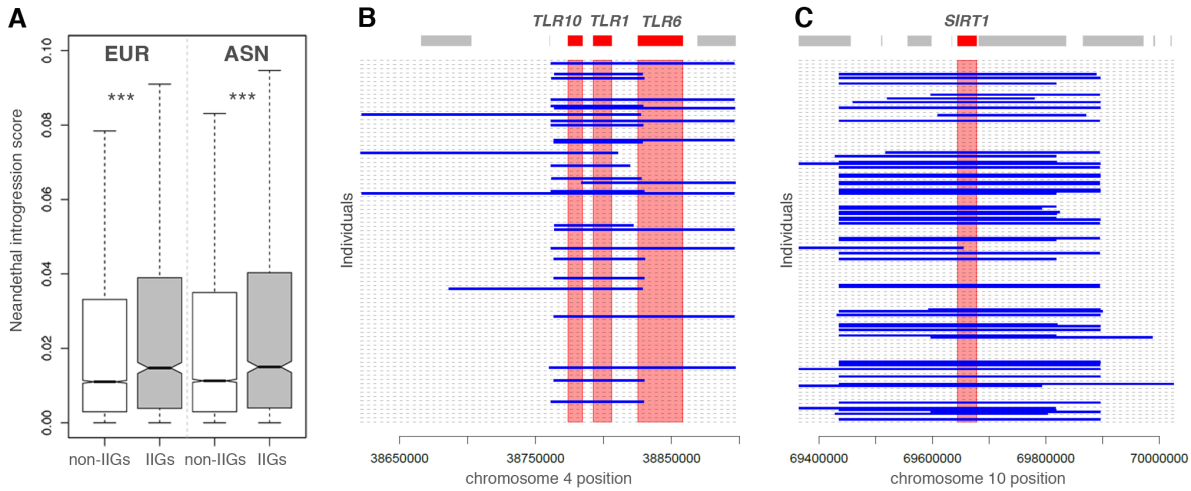

A
f value
non-IIGs
IIGs
B
Odds ratio
Percentile of the f distribution (%)
C
f value
Sensors
Adaptors
Signal transducers
Transcription
Effectors
Secondary receptors
Accesory molecules
Regulators
UC
D

Figure 3


A
TLR10
TLR1
TLR6
Proportion of FCS outliers
FCS values
chromosome 4 position
B
IFIH1
chromosome 2 position
C
MERTK
chromosome 2 position
D
ZFPM2
chromosome 8 position

1 Figure 4

2


A
EUR
ASN
***
***
Neandethal introgression score
0.00
0.02
0.04
0.06
0.08
0.10
non-IIGs
IIGs
non-IIGs
IIGs
B
TLR10
TLR1
TLR6
Individuals
38650000
38750000
38850000
chromosome 4 position
C
SIRT1
Individuals
69400000
69600000
69800000
70000000
chromosome 10 position

3

## 1.3. Résumé des travaux et discussion spécifique

Bien que de nombreuses études aient identifié les forces de sélection ayant guidé l'évolution de certains gènes impliqués dans l'immunité innée ainsi que l'intensité de l'action de ces forces, toutes ont utilisé une approche basée sur des gènes candidats. Dans cette étude, nous avons pour la première fois effectué une analyse des pressions de sélection exercées sur l'ensemble des gènes de l'immunité et montré que la diversité génétique présente dans ces gènes résulte de l'action d'évènements complexes. Ainsi, nous avons dans un premier temps montré que les gènes de l'immunité font partie des séquences codantes les plus contraintes du génome, soulignant leur importance biologique. Cependant, l'intensité de l'action de la sélection purificatrice n'est pas homogène parmi les classes de gènes de l'immunité innée, ni même au sein de ces classes. Nous avons ainsi montré que les contraintes sélectives exercées sur les gènes codant pour des molécules impliquées dans la reconnaissance des pathogènes sont très variables alors que les molécules impliquées dans la transduction du signal et l'initiation et la propagation de la réponse transcriptionnelle sont toutes particulièrement contraintes.

Nous avons par la suite développé une méthode composite puissante pour détecter les signatures de sélection positive dans le génome et montré que les gènes de l'immunité innée n'ont pas été préférentiellement ciblés par ce type de sélection. Cependant, nous avons mis en évidence un certain nombre de gènes impliqués dans la réponse immunitaire innée présentant des signes d'adaptation des populations humaines à leur environnement. Nous avons par la suite identifié un certain nombre de variations génétiques qui pourraient être à l'origine des évènements de sélection positive.

Finalement, nous avons mis à profit les données récemment rendues publiques d'estimation de l'introgression de séquences provenant de Néandertal dans le génome humain pour décrire la proportion d'allèles présents dans les gènes de l'immunité innée provenant d'Hommes archaïques. Nous avons ainsi montré que l'intensité de la sélection négative exercée sur les variations génétiques provenant de Néandertal est plus faible dans les gènes de l'immunité innée que dans le reste des séquences codantes du génome. Nous avons par la suite identifié des gènes présentant un taux d'introgression parmi les plus élevés des séquences codantes. De manière particulièrement intéressante, nous avons trouvé parmi ces gènes le cluster *TLR6-1-10* que nous avons confirmé au cours de cette étude comme étant sous sélection positive dans les populations européennes. Un autre gène présente les mêmes

caractéristiques dans les populations d'Asie de l'Est : *SIRT1*. Ces deux régions génomiques pourraient donc avoir participé à la différenciation des populations suite aux mélanges génétiques qui se sont faits avec l'Homme de Néandertal. Des études plus poussées permettraient de confirmer ou d'infirmer l'hypothèse selon laquelle ces régions ont évolué selon des évènements d'introgression adaptative.

Notre étude permet, pour la première fois, d'apprécier l'hétérogénéité et la complexité des pressions de sélection ayant agi sur les gènes de l'immunité innée pris dans leur ensemble. Dans un premier temps, notre analyse de l'action de la sélection purificatrice permet de proposer une classification de l'importance fonctionnelle relative de chacun des gènes étudiés. En effet, les séquences n'ayant accumulé que peu de mutations non-synonymes semblent avoir une fonction biologique capitale. À titre d'exemples, nous trouvons que *STAT1* et *TRAF3* ne sont que très peu permissifs à l'accumulation de mutations non-synonymes. Des mutations dans ces gènes ont été décrites comme causant des IP (Casanova et al., 2013), montrant qu'ils sont essentiels au bon déroulement de la réponse immunitaire. Dans un second temps, notre analyse par catégories nous permet d'émettre l'hypothèse que certains mécanismes de reconnaissance des pathogènes, notamment l'identification de leurs acides nucléiques, sont essentiels à la défense de l'hôte alors que d'autres semblent redondants.

Nous avons ensuite développé une méthode puissante associant les résultats de différentes statistiques permettant de détecter l'action de la sélection positive sur des séquences pour identifier des gènes ayant participé à l'adaptation spécifiques de populations. La datation des évènements de sélection positive que nous avons identifiés se révèle être en accord avec l'hypothèse selon laquelle la première révolution agricole aurait induit une profonde modification de l'exposition de l'Homme face aux pathogènes, ce qui aurait mené à une adaptation des populations à leur environnement particulièrement marquée. Cependant, nous avons aussi identifié des évènements de sélection particulièrement récents. Ainsi, nous avons pour la première fois montré que l'évènement de sélection sur la mutation 1264G du gène *CD36*, qui a été aussi rapportée par plusieurs autres études comme étant sous sélection positive chez les Yorubas, remonte à 3 600 ans. Cela supporte l'idée que cet évènement de sélection est particulièrement fort et récent.

Finalement, nous avons pour la première fois estimé la part de diversité génétique présente dans les gènes de l'immunité que nous aurions pu hériter de l'Homme de

Néandertal. Nos résultats montrent que l'introgression d'allèles de ces hominidés aujourd'hui éteints n'est pas homogène parmi les gènes de l'immunité innée. Certains récepteurs tels que les gènes *OAS* ou le cluster *TLR6-1-10* montrent de fortes probabilités d'ascendance néandertalienne ainsi que des preuves moléculaires notées dans cette étude ou dans d'autres d'évolution sous sélection positive. Cela nous laisse supposer que l'apport de diversité génétique supplémentaire dans ces régions géniques a pu être avantageux pour *Homo sapiens*, lui permettant d'acquérir des innovations moléculaires adaptées au nouvel environnement auquel il était confronté lors de sa cohabitation avec Néanderthal.

Notre étude prise dans son ensemble améliore les connaissances que nous avons de l'importance et la redondance des gènes de l'immunité innée et apporte donc un éclairage nouveau sur la participation de l'Homme de Néandertal à la diversité génétique retrouvée dans ces gènes.

# 2. Caractérisation du contrôle génétique de l'expression des miARN au cours de l'infection par *Mycobacterium tuberculosis*

## 2.1. Contexte

Comme nous l'avons évoqué au cours de l'introduction, la susceptibilité aux maladies, et en particulier aux maladies infectieuses, est variable entre les individus. Ce phénotype macroscopique résultant de l'action de multiples facteurs, il est excessivement compliqué à étudier dans son ensemble. C'est la raison pour laquelle, au cours de ces dernières années, les phénotypes cellulaires et moléculaires sous-jacents ont été utilisés comme phénotypes intermédiaires permettant de mieux comprendre les différences de réponses immunitaires observées entre individus. Ces phénotypes présentent l'avantage de pouvoir être étudiés *ex vivo* et le développement de nombreux outils de biologie cellulaire, de biologie moléculaire et de mesure de l'expression génique à large échelle permettent aujourd'hui de les examiner en détail. La réponse immunitaire innée met en jeu de nombreux acteurs cellulaires. Parmi ceux-ci, les cellules dendritiques jouent un rôle clé dans l'initiation de la réponse adaptative. En

effet, ces cellules sont, comme les macrophages, des cellules présentatrices d'antigènes. Elles ont pour fonction de phagocyter les pathogènes puis de migrer vers les ganglions lymphatiques où elles présenteront les antigènes aux lymphocytes B et T, engageant par la suite la réponse très spécifique de ces cellules de l'immunité adaptative. Ainsi, l'intensité de leur activation et de leur réponse au stimulus aura une conséquence importante sur le devenir de l'infection. Outre des changements d'aspect très particulier, avec l'apparition de dendrites, l'activation des cellules dendritiques par la phagocytose de particules antigéniques ou la réception de signaux de danger s'accompagne de nombreuses modifications d'expression de gènes qui varient entre individus. De récentes études ont démontré que des variations génétiques pouvaient affecter les niveaux d'expression de certains ARNm dans ces cellules suite à l'infection par *Mycobacterium tuberculosis*, illustrant l'interaction entre facteurs génétiques et environnementaux qui est à la base de la variabilité interindividuelle de susceptibilité aux maladies infectieuses. Les miARN ont été largement décrits au cours de ces dernières années comme des acteurs majeurs de la régulation de l'expression des ARNm impliqués dans de très nombreux processus cellulaires. L'expression des miARN est très variable entre types cellulaires, entre tissus, entre états d'activation des cellules ainsi qu'entre individus. De plus, des analyses ont révélé qu'ils pourraient permettre d'expliquer une partie des variations interindividuelles. Si l'implication des miARN dans la régulation des réponses immunitaires de l'hôte est de plus en plus largement décrite, aucune étude n'a pour le moment été faite sur les variabilités interindividuelles d'expression de ces petits ARN non-codants suite à l'infection.

Le premier objectif de notre étude était donc de caractériser les modifications d'expression des miARN induites dans les cellules dendritiques suite à une stimulation infectieuse : l'infection par *Mycobacterium tuberculosis*. Nous avons ensuite souhaité caractériser le contrôle génétique de ces différences d'expression. Finalement, nous avons utilisé une combinaison d'approches bioinformatiques et expérimentales pour décrire les régulations de l'expression des ARNm et de la production de cytokines des cellules dendritiques suite à l'infection en nous appuyant plus particulièrement sur l'exemple de miR-29.

## 2.2. Article

Research

# A genomic portrait of the genetic architecture and regulatory impact of microRNA expression in response to infection

Katherine J. Siddle,[1,2,11] Matthieu Deschamps,[1,2,3,11] Ludovic Tailleux,[4] Yohann Nédélec,[5,6] Julien Pothlichet,[1,2] Geanncarlo Lugo-Villarino,[7,8] Valentina Libri,[9] Brigitte Gicquel,[4] Olivier Neyrolles,[7,8] Guillaume Laval,[1,2] Etienne Patin,[1,2] Luis B. Barreiro,[6,10] and Lluís Quintana-Murci[1,2,12]

*[1]Institut Pasteur, Unit of Human Evolutionary Genetics, 75015 Paris, France; [2]Centre National de la Recherche Scientifique, CNRS URA3012, 75015 Paris, France; [3]Université Pierre et Marie Curie, Cellule Pasteur UPMC, 75015 Paris, France; [4]Institut Pasteur, Unit of Mycobacterial Genetics, 75015 Paris, France; [5]Department of Biochemistry, Faculty of Medicine, University of Montréal, Montréal H3T 1C5, Canada; [6]Ste-Justine Hospital Research Centre, Montréal H3T 1C5, Canada; [7]Centre National de la Recherche Scientifique, Institut de Pharmacologie et de Biologie Structurale, 31077 Toulouse, France; [8]Université de Toulouse, Université Paul Sabatier, Institut de Pharmacologie et de Biologie Structurale, 31077 Toulouse, France; [9]Institut Pasteur, Centre d'Immunologie Humaine, 75015 Paris, France; [10]Department of Paediatrics, Faculty of Medicine, University of Montréal, Montréal H3T 1C5, Canada*

**MicroRNAs (miRNAs) are critical regulators of gene expression, and their role in a wide variety of biological processes, including host antimicrobial defense, is increasingly well described. Consistent with their diverse functional effects, miRNA expression is highly context dependent and shows marked changes upon cellular activation. However, the genetic control of miRNA expression in response to external stimuli and the impact of such perturbations on miRNA-mediated regulatory networks at the population level remain to be determined. Here we assessed changes in miRNA expression upon *Mycobacterium tuberculosis* infection and mapped expression quantitative trait loci (eQTL) in dendritic cells from a panel of healthy individuals. Genome-wide expression profiling revealed that ~40% of miRNAs are differentially expressed upon infection. We find that the expression of 3% of miRNAs is controlled by proximate genetic factors, which are enriched in a promoter-specific histone modification associated with active transcription. Notably, we identify two infection-specific response eQTLs, for miR-326 and miR-1260, providing an initial assessment of the impact of genotype-environment interactions on miRNA molecular phenotypes. Furthermore, we show that infection coincides with a marked remodeling of the genome-wide relationships between miRNA and mRNA expression levels. This observation, supplemented by experimental data using the model of miR-29a, sheds light on the role of a set of miRNAs in cellular responses to infection. Collectively, this study increases our understanding of the genetic architecture of miRNA expression in response to infection, and highlights the wide-reaching impact of altering miRNA expression on the transcriptional landscape of a cell.**

**[Supplemental material is available for this article.]**

The responses of host immune cells to microbial stimuli are accompanied by marked changes in gene expression, with transcriptional programs that can be shared among different microbes or be specific to each (Huang et al. 2001; Amit et al. 2009; Chevrier et al. 2011; Gat-Viks et al. 2013). The regulatory networks that control the initiation, peak magnitude, and resolution of host responses must all be properly tuned to achieve optimal immunity. In this context, microRNAs (miRNAs), a group of endogenous small RNAs (~22 nt), play a critical role in the epigenetic regulation of gene expression in eukaryotes (Ambros 2004; Bartel 2004). MiRNAs bind complementary sequences of target mRNAs to promote translational repression and/or mRNA degradation (Guo et al. 2010; Huntzinger and Izaurralde 2011). For an individual target, miRNAs have only subtle regulatory effects (Hornstein and Shomron 2006; Baek et al. 2008; Selbach et al. 2008), though a single miRNA may target over 100 genes. Over 60% of human genes are expected to be directly regulated by miRNAs (Friedman et al. 2009), with many predicted to be cotargeted by multiple miRNAs (Krek et al. 2005; Stark et al. 2005; Tsang et al. 2010). The importance of such complex and tightly regulated miRNA–mRNA networks is highlighted by the strong evolutionary constraints acting on both miRNAs and mRNA target sites, across species and within humans (Chen and Rajewsky 2006; Saunders et al. 2007; Friedman et al. 2009; Quach et al. 2009; Christodoulou et al. 2010; Berezikov 2011).

In addition to their involvement in a wide range of biological processes, including development, cell differentiation, and apo-

**[11]These authors contributed equally to this work.**
**[12]Corresponding author**
**E-mail quintana@pasteur.fr**
Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.161471.113.




24:850–859 Published by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/14; www.genome.org

ptosis, the role of miRNAs in regulating mammalian immune systems is increasingly well established (Lodish et al. 2008; O'Connell et al. 2012). MiRNAs regulate the development and function of cells of innate and adaptive immunity (Chen et al. 2004; Johnnidis et al. 2008; Lodish et al. 2008; O'Connell et al. 2012), and can have either pro-inflammatory or anti-inflammatory effects, indicating that the immune system utilizes multiple miRNAs to balance its response (O'Connell et al. 2012). Furthermore, experimental data indicate that viral, parasitic, and bacterial pathogens induce marked changes in miRNA expression in host cells (Cullen 2011; Eulalio et al. 2012). For example, activation of the innate immunity Toll-like receptor (TLR) pathway influences the expression of a well-defined group of miRNAs, while miRNAs can also regulate TLR expression as well as that of TLR signaling molecules, transcription factors, and cytokines (O'Neill et al. 2011).

There is growing evidence indicating that there is strong variation in miRNA expression in human populations, within a given cellular state, cell type, or tissue (Wang et al. 2009; Huang et al. 2011; Lu and Clark 2012; Lappalainen et al. 2013). The extent to which this variation is under genetic control (i.e., miRNA expression quantitative trait loci, miR-eQTLs) has recently begun to be investigated (Borel et al. 2011; Rantalainen et al. 2011; Gamazon et al. 2012; Parts et al. 2012; Civelek et al. 2013; Gamazon et al. 2013; Lappalainen et al. 2013). However, as has been shown for mRNAs in yeast and mammals, genetic variants can differentially affect gene expression after perturbation by various treatments or environmental variables (i.e., response eQTLs) (Gargalovic et al. 2006; Smith and Kruglyak 2008; Smirnov et al. 2009; Yang et al. 2009; Dombroski et al. 2010; Romanoski et al. 2010; Maranville et al. 2011; Barreiro et al. 2012). In humans, recent studies of protein-coding gene expression have identified response eQTLs related to oxidative stress (Romanoski et al. 2010), ionizing radiation (Smirnov et al. 2009), drug treatment (Maranville et al. 2011), and infection (Barreiro et al. 2012; Idaghdour et al. 2012). Conversely, as the few miR-eQTL studies to date have all used steady-state expression measurements, the degree of population variation in miRNA expression upon external stimulation, and the extent to which gene-environment interactions may affect the regulation of miRNA responses, remain to be determined.

Here, we aimed to dissect the genetic architecture and regulatory impact of miRNA expression in response to an external, infectious stimulus. To do so, we first characterized the population variation of miRNA transcriptional responses to infection using, as a model, *Mycobacterium tuberculosis* (MTB) infection of human dendritic cells (DCs). We then investigated the extent to which miRNA expression variation upon infection is under genetic control, providing the first attempt to map response miR-eQTLs. We next explored the relationship between miRNA and mRNA expression levels to understand how infection not only affects miRNA responses but also impacts upon broader miRNA-mediated regulatory networks. Finally, we performed miRNA gain- and loss-of-function experiments to assess the impact of altered miRNA expression on downstream transcriptional and protein responses to infection.

## Results

### Genomic characterization of miRNA transcriptional responses to infection

We profiled genome-wide patterns of miRNA expression in monocyte-derived DCs, untreated and after infection with a virulent strain of MTB, from a panel of 65 healthy individuals of European descent. The presence of an infection-related response in these samples has been previously confirmed at the mRNA level, by the altered expression of genes involved in immune responses, and at the protein level, by the strong induction of cytokines known to play a critical role in protective immunity against tuberculosis (TB) (Barreiro et al. 2012). After quality checks and normalization of the data, we assessed differences in miRNA expression levels upon infection using a final set of 346 miRNAs from 63 individuals (Supplemental Fig. S1; Supplemental Methods). We found that 155 miRNAs were differentially expressed upon infection ($P < 1 \times 10^{-5}$; Bonferroni-corrected $P < 0.01$), of which 64 were up-regulated and 91 down-regulated (Fig. 1; Supplemental Table S1). Among these, down-regulated miRNAs exhibited lower fold changes than those that were up-regulated, with only three (3%), compared with 20 (31%), showing at least a twofold difference in expression levels upon infection (Fig. 1). These maximal fold changes are markedly smaller than those observed for protein-coding genes in the same system (Barreiro et al. 2012), consistent with previous observations (e.g., Sharbati et al. 2011).

The most differentially expressed miRNAs upon infection include, among others, the up-regulated miR-155, miR-132, and miR-29a, and the down-regulated miR-361-5p, miR-185, and miR-27a. These miRNAs are involved in the modulation of immune functions, such as the activation of core signaling pathways or the response to bacterial infections (O'Neill et al. 2011; Eulalio et al. 2012; Qi et al. 2012). More generally, although we found a substantial overlap between the 40% of differentially expressed miRNAs in our study and previous analyses using similar settings (Ceppi et al. 2009; Liu et al. 2009; Fu et al. 2011; Sharbati et al. 2011; Maertzdorf et al. 2012; Yi et al. 2012), we also identified distinctive signatures for some miRNAs in our model. These include the down-regulation of miR-125b and members of the miR-148

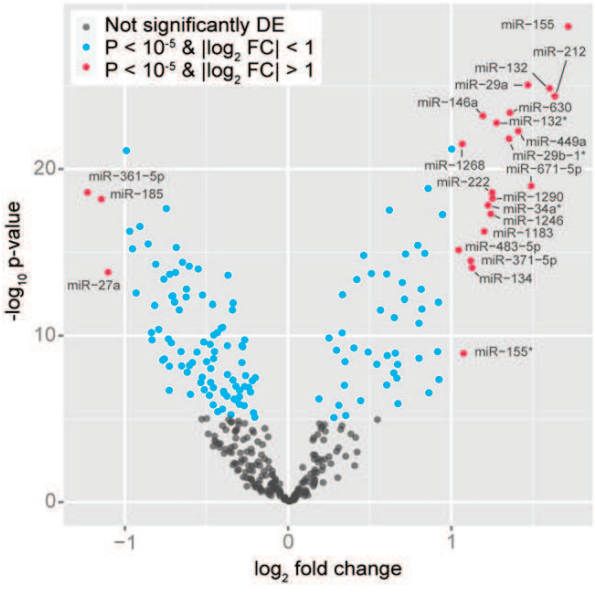



**Figure 1.** Changes in miRNA expression levels upon infection. Volcano plot showing the differential expression of miRNAs in DCs upon infection with MTB. Red dots denote significantly differentially expressed miRNAs whose expression changed by more than twofold following infection, whereas blue dots represent significantly differentially expressed miRNAs whose fold change (FC) was less than two. (DE) Differentially expressed.

family, which have been reported to be up-regulated upon MTB infection of macrophages (Rajaram et al. 2011) and activation of DCs with lipopolysaccharide (Liu et al. 2010), respectively. Additionally, we identified previously unreported miRNAs, such as miR-630 and miR-339-3p, as being differentially expressed upon infection. Collectively, our results are consistent with a general pro-inflammatory response (Turner et al. 2011), together with putatively cell or stimulus-dependent responses that could impact the ways in which infection is established and maintained.

## Genetic regulation of miRNA expression upon infection

To identify genetic variants that affect the response to MTB infection, we mapped miR-eQTLs by testing the association between miRNA expression profiles and genome-wide genotyping data from the same 63 individuals (Methods). This sample size affords sufficient power to detect eQTLs, even in some cases where genetic variation has a moderate impact on expression levels (Supplemental Fig. S2; Supplemental Methods). To map *cis*-acting variants, we focused on SNPs located within a 200-kb window centered on the mature miRNA, and analyzed the data separately for non-infected and infected samples to discern miR-eQTLs that are shared between conditions from those that are unique to a particular state. In total, we identified miR-eQTLs for six miRNAs (miR-326, miR-338-3p, miR-451, miR-1260, miR-769-5p, and miR-130b) in infected samples, of which one (miR-338-3p) was also significant in non-infected samples, at a false discovery rate (FDR) of 0.2 (Table 1; Supplemental Table S2). These associations accounted for 20%–50% of the variance in the expression of these miRNAs.

Despite the fact that most of these miR-eQTLs displayed a significant association only in infected samples, we observed similar tendencies in the effect of the genotype on miRNA expression before and after infection (Supplemental Fig. S3). To identify response eQTLs, where a genetic variant has a stimulus-specific impact on transcript abundance, we focused on miR-eQTLs detected in one condition that had no observed effect in the other condition ($P > 0.05$). Using this threshold, we detected two putative response miR-eQTLs for miR-326 and miR-1260 (Fig. 2A). Using BRIdGE, a Bayesian approach for identifying genetic associations under different models of gene-environment interactions (Maranville et al. 2011), we confirmed five of the six associations detected, including the two response miR-eQTLs, at a posterior probability >0.7 (FDR = 0.15; Supplemental Table S3). This approach further enabled us to identify a general interaction effect for miR-338-3p, where the miR-eQTL had a different effect in each condition (Supplemental Fig. S3).

**Table 1.** miR-eQTLs identified in *cis* for miRNA expression variation in non-infected and/or MTB-infected samples

| | Non-infected samples | | Infected samples | |
|---|---|---|---|---|
| miRNA | SNP[a] | Minimum *P*-value[b] | SNP[a] | Minimum *P*-value[b] |
| miR-326 | rs658573 | $6.39 \times 10^{-2}$ | rs532751 | $5.15 \times 10^{-6}$ |
| miR-338-3p[c] | rs4969258 | $1.18 \times 10^{-10}$ | rs7220048 | $9.21 \times 10^{-6}$ |
| miR-451 | rs9279[d] | $1.49 \times 10^{-2}$ | rs9279[d] | $1.45 \times 10^{-5}$ |
| miR-1260 | rs4899651 | $7.36 \times 10^{-2}$ | rs11159250 | $8.35 \times 10^{-5}$ |
| miR-769-5p | rs759623 | $2.40 \times 10^{-2}$ | rs8111976 | $1.75 \times 10^{-4}$ |
| miR-130b | rs3788329 | $7.63 \times 10^{-4}$ | rs3788329 | $2.09 \times 10^{-4}$ |

[a]SNP for which the strongest association with miRNA expression was observed in a given condition.
[b]Significance was determined using an FDR <0.2 ($P < 4.04 \times 10^{-5}$ and $2.31 \times 10^{-4}$ for non-infected and infected samples, respectively).
[c]A significant eQTL was detected in both non-infected and infected conditions for this miRNA.
[d]More than one SNP showing the same minimum *P*-value for association with miRNA expression. See Supplemental Table S2 for full list of associated SNPs.

We next searched for *trans*-eQTLs by associating miRNA expression levels with genome-wide SNPs. We identified one miRNA, miR-582-5p, whose expression was associated with a cluster of SNPs in infected samples (Fig. 2B), of which the most strongly associated was rs12523473 (Bonferroni-corrected $P = 1.49 \times 10^{-4}$). While this association lies outside the 200-kb region considered for *cis*-acting variants, its physical proximity to miR-582-5p suggests that this miR-eQTL is likely to be a long-distance *cis*-eQTL. The detection of *trans*-eQTLs suffers, however, from a high burden of multiple testing while effects may be only of modest size (Gilad et al. 2008; Majewski and Pastinen 2011; Montgomery and Dermitzakis 2011). Given that regulatory variants have been observed to overlap with SNPs associated with complex phenotypes (Nica et al. 2010; Nicolae et al. 2010), we restricted our analysis to SNPs that have been suggestively associated ($P < 1 \times 10^{-5}$) with TB susceptibility by genome-wide association studies (GWAs) (Hindorff et al. 2009; Thye et al. 2010; Thye et al. 2012). In doing so, we identified a putative *trans*-eQTL for let-7i, a strongly induced, pro-inflammatory miRNA, the expression of which was associated with rs9373523 upon MTB infection (Bonferroni-corrected $P = 7.78 \times 10^{-3}$). The mechanism underlying this association remains unclear, however, that this SNP lies in an intron of the gene *STXBP5*, which is a target of let-7i, may point to a complex regulatory interaction between these transcripts.

## Genomic and functional context of miR-eQTLs

To understand how eQTLs may influence miRNA expression, we studied the genomic context of the detected miR-eQTLs. We first investigated the overlap between miR-eQTLs and signatures of regulatory regions, ChIP-seq peaks and DNase I signals, reported by the ENCODE Project for human monocytes, the closest available cell-type to our model (The ENCODE Project Consortium 2012). We observed a number of overlaps with histone modifications (Supplemental Table S4). In particular, we found that the detected miR-eQTLs are significantly enriched in regions associated with the histone modification H3K4me3 ($P = 1.6 \times 10^{-3}$). That this is a mark of regulatory elements primarily associated with promoters/transcription start sites (The ENCODE Project Consortium 2012), and that three of the four SNPs overlapping such regions are located 5′ of the miRNA with which they are associated (miR-1260, miR-582-5p, and miR-769-5p), further supports the functional relevance of the miR-eQTL regions in modulating miRNA expression.

To gain further insight into the putative regulatory mechanisms underlying the detected miR-eQTLs, we refined our association signals by genotype imputation (Supplemental Fig. S4; Supplemental Methods). We then searched for our miR-eQTL SNPs, and those in strong linkage disequilibrium with them, among (1) DNase I sensitivity QTLs (dsQTLs) (Degner et al. 2012), (2) mRNA-eQTLs identified in the same cellular infection model (Barreiro et al. 2012), and more generally, (3) mRNA-eQTLs identified from the International HapMap project (Xia et al. 2012). We found that two miR-eQTL SNPs for miR-451 were associated with variation in the chromatin accessibility of a nearby region (rs9279, $P = 3.77 \times 10^{-10}$, and rs2320588, $P = 9.55 \times 10^{-8}$). This DNase I-sensitive region, which lies 4 kb upstream of the precursor of miR-

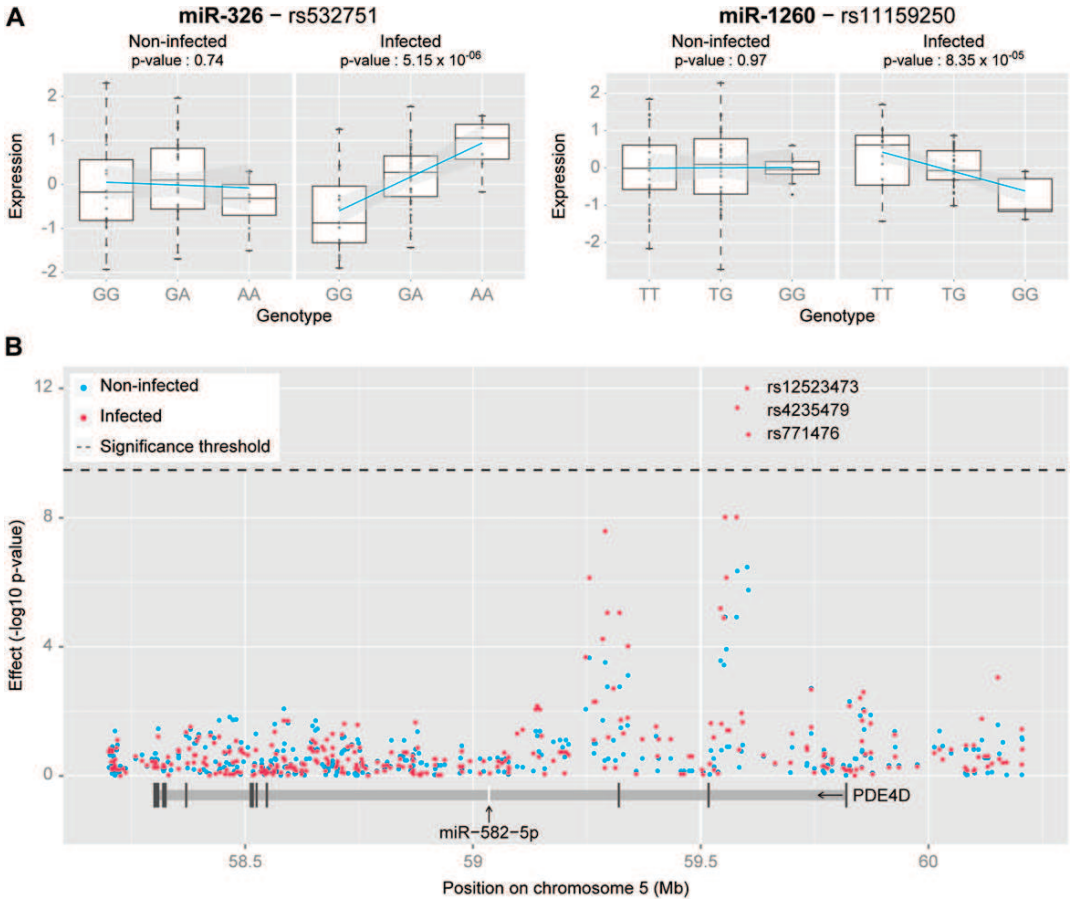



**Figure 2.** MiR-eQTLs upon infection. (*A*) Boxplots showing the detected response miR-eQTLs in *cis* for miR-326 and miR-1260, in non-infected and infected samples. (*B*) Regional association plot for miR-582-5p and genotyped SNPs in the region of the gene *PDE4D* showing the location of a cluster of significantly associated SNPs ~500 kb upstream of the miRNA in infected samples. An additional region, in between the detected eQTL and the miRNA, also showed a strong tendency of association; however, this did not reach genome-wide significance. All annotations are based on UCSC hg19.

451, is further associated with the binding of a number of transcription factors, especially POLR2A and GATA1 (The ENCODE Project Consortium 2012), suggesting that this region may directly regulate the transcription of miR-451. With respect to mRNAs, miR-eQTL SNPs for miR-582-5p were associated with the expression of *DEPDC1B* in *cis* (FDR = 0.05–0.11) and *PNMAL1* in *trans* (FDR < 0.2), the latter gene being a predicted target of miR-582-5p, in HapMap samples. Interestingly, four of the reported miR-eQTLs (miR-326, miR-338-3p, miR-130b, and miR-582-5p) are non-canonical intronic miRNAs (mirtrons) (Okamura et al. 2007; Ruby et al. 2007). However, no *cis*-eQTLs were identified for the corresponding host genes (*ARRB1*, *AATK*, *PPIL2*, and *PDE4D*, respectively) in the same samples. This, together with the observation that only miR-582-5p expression was moderately positively correlated with that of its host gene ($r = 0.311$; $P = 0.013$), supports the notion that mirtrons are regulated and/or processed independently of their host genes (Parts et al. 2012; Civelek et al. 2013).

## Extensive remodeling of genome-wide miRNA–mRNA interactions upon infection

To understand the impact of infection on broader miRNA-mediated regulatory networks, we next investigated the genome-wide relationships between variation in miRNA expression and that of protein-coding genes. To do so, we calculated Pearson correlation coefficients between the expression levels of mature miRNAs and mRNAs, obtained from the same 63 individuals, before and after infection (see Methods). We detected an overwhelming 35-fold increase in the number of significant correlations between miRNAs and mRNAs in infected samples with respect to non-infected samples (FDR < 0.005; Supplemental Table S5; Supplemental Fig. S5). Furthermore, the patterns of genome-wide correlations strongly differed between the two conditions: Correlations before infection were enriched in positive correlations ($P = 6 \times 10^{-4}$) whereas those after infection were skewed toward negative relationships ($P < 1 \times 10^{-20}$), compared with the null distribution. Specifically, only 23% of significant miRNA–mRNA correlations were negative in non-infected samples, compared with 52% in infected samples (Supplemental Fig. S6), a difference that became more pronounced when considering only the strongest correlations ($|r| > 0.7$, 9% vs. 72% in non-infected and infected samples, respectively). These trends remained similar after accounting for variation in the percentage of infected cells among individuals (Supplemental Fig. S7). Notably, we observed an enrichment in predicted miRNA targets not only among negatively ($P = 2.85 \times 10^{-5}$) but also among positively ($P = 6.45 \times 10^{-7}$) correlated genes

in infected cells. This observation is consistent with the fact that miRNAs are often involved in incoherent feed-forward loops and other complex network relationships with their targets (Hornstein and Shomron 2006; Tsang et al. 2007; Vasudevan et al. 2007; Martinez et al. 2008; Ebert and Sharp 2012; Lu and Clark 2012).

The majority of significant miRNA–mRNA correlations in both conditions were accounted for by a small number of miRNAs, as previously observed (Nunez-Iglesias et al. 2010; Su et al. 2011; Gamazon et al. 2012). Upon infection, 46 miRNAs were each correlated with at least 10 mRNAs, of which 15 were found to be associated with over 100 mRNAs and cumulatively accounted for 75% of all significant correlations (Fig. 3A; Supplemental Table S5). Furthermore, these 46 miRNAs were enriched in differentially expressed miRNAs upon infection ($P = 0.03$; $N = 30$), and included many of known importance in the regulation of the immune response (e.g., miR-155, miR-132, and miR-146a). These differentially expressed miRNAs and their significantly correlated gene sets, which were also enriched in differentially expressed genes ($P < 1 \times 10^{-20}$), formed a tightly connected regulatory network (Fig. 3B). This highlights the highly interrelated nature of miRNAs in gene regulation, consistent with their frequent cotargeting of mRNA transcripts (Krek et al. 2005; Stark et al. 2005; Tsang et al. 2010). Furthermore, among these correlated gene sets, 70% of those presenting an over/under-representation of at least one KEGG pathway and/or GO category (FDR<0.05) were associated with immune-related functions. These included innate immunity signaling pathways (e.g., TLR and JAK-STAT pathways) and activation and differentiation of lymphocytes (Supplemental Table S6), consistent with the expected functions of activated DCs. Taken together, these findings suggest that a subset of differentially expressed miRNAs may account for most of the functional associations between miRNAs and mRNAs upon MTB infection.

## Characterization of the impact of miR-29a on the response to infection

To experimentally assess the impact of altered miRNA expression on both miRNA–mRNA interactions and cellular responses to MTB infection, we studied one miRNA—miR-29a—using gain- and loss-of-function approaches. We chose this miRNA because (1) it is strongly induced upon infection (Fig. 1), (2) it is among those presenting the largest number of correlated mRNAs in infected cells (Fig. 3), and (3) it has been explicitly implicated in the response to various mycobacterial infections (Fu et al. 2011; Ma et al. 2011; Sharbati et al. 2011; Eulalio et al. 2012; Yi et al. 2012; Brain et al. 2013). We thus transfected DCs with a miR-29a mimic or a miR-29 family inhibitor, as all miR-29 family members (miR-29a,b,c) share the same seed sequence, and confirmed the perturbation of miR-29 expression before and after infection with MTB (Supplemental Fig. S8; Supplemental Methods). Using genome-wide expression arrays, we identified 193 and 539 differentially expressed genes, at the steady-state, and 59 and 307, after infection, in miR-29 overexpressing and inhibited cells, respectively, compared with control-transfected cells (FDR-corrected $P < 0.01$,

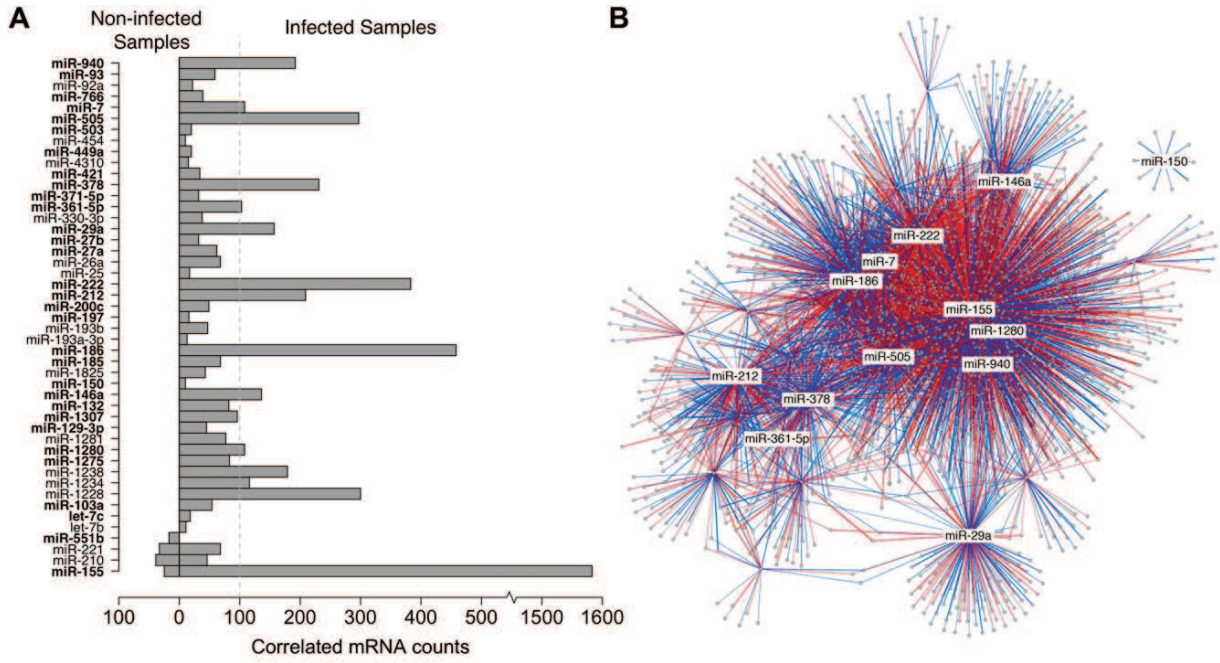



**Figure 3.** Relationship between the levels of expression of miRNAs and protein-coding genes. (*A*) Barplot showing the number of significantly correlated mRNAs per miRNA in non-infected and infected samples. Only miRNAs whose expression levels were significantly correlated with those of at least 10 mRNAs are shown. A total of 47 miRNAs were correlated with at least 10 mRNAs in non-infected and/or infected conditions, with three miRNAs satisfying this criterion in both conditions. A total of 31 of these 47 miRNAs were significantly differentially expressed upon MTB infection (marked in bold). (*B*) Regulatory network of significantly correlated mRNAs and differentially expressed miRNAs in MTB-infected samples. Nodes represent miRNAs and mRNAs. MiRNAs are labeled when correlated with more than 100 mRNAs, with the exception of miR-150, which is independent of the main network. Edge thickness reflects the strength of the correlation between one miRNA and one mRNA transcript. Edge color represents the direction of the correlation (red, negative; blue, positive).

Supplemental Table S7). These differentially expressed genes, in particular upon miR-29a overexpression, were enriched in predicted targets of miR-29a ($P = 0.01–1 \times 10^{-20}$).

Importantly, we found that the genes whose expression was significantly correlated with that of miR-29a after infection in our previous computational analysis (Fig. 3; Supplemental Table S5) showed significantly greater changes in their expression levels with respect to noncorrelated genes in infected samples ($P = 2.31 \times 10^{-5}$ and $P = 9.34 \times 10^{-3}$ in overexpressing and inhibited cells, respectively) (Fig. 4A,B). Moreover, miR-29a predicted targets whose expression was correlated with that of this miRNA showed a significant decrease and increase in their expression levels in overexpression and inhibition experiments, respectively (Fig. 4C,D). These results support that a significant proportion of miRNA-correlated transcripts, in particular those that are predicted to be direct targets, are indeed causally regulated by miR-29.

Lastly, we assessed the impact of miR-29a on cellular responses to infection at the transcript and protein levels. Consistent with the profound effect of infection on DC maturation and function, the transcriptional response to MTB was highly concordant between control- and miR-29-transfected cells (82% and 91% overlap in mimic and inhibitor-transfected cells, respectively; FDR-corrected $P < 0.01$; Supplemental Table S7). However, a subset of differentially expressed genes upon infection responded differently following perturbation of miR-29 (64 and 235 in mimic and inhibitor-transfected cells, respectively; $P < 0.01$; Supplemental Table S7). Among these differentially responding genes, we observed an enrichment of a number of KEGG pathways and GO categories (Supplemental Table S8). In particular, genes that were differentially up-regulated following miR-29 inhibition were significantly enriched in genes involved in cytokine–cytokine receptor interaction and TLR signaling pathways. At the protein level, we similarly observed a major effect of infection, with a strong induction of many cytokines that play a major role in the DC response to infection, including TNF, IL12B, and IL10 (Giacomini et al. 2001; Hickman et al. 2002; Supplemental Table S9). However, the inhibition of miR-29 promoted an enhanced cytokine response, as attested to by the significantly higher induction of 12 cytokines and chemokines, with respect to control-transfected samples, after infection (Fig. 5A). Notably, three of these cytokines are miR-29a predicted targets (*IL12B, IL2RA*, and *CXCL10*), consistent with a direct effect of miR-29 on cytokine responses. Furthermore, despite the more modest effect of miR-29a overexpression on cytokine levels (Supplemental Table S9), one of these proteins, CXCL10, was significantly down-regulated compared with control-transfected samples (Fig. 5B), providing strong support for a direct regulation of this chemokine by miR-29 in MTB-infected DCs.

## Discussion

Identifying genetic variants that affect miRNA expression in the presence or absence of specific environmental variables can provide insights into the mechanisms underlying variation in transcript abundance. We first showed that the expression of 3% of miRNAs is explained by proximate genetic factors, consistent with several previous estimates (Rantalainen et al. 2011; Parts et al. 2012; Civelek et al. 2013). That 9% of protein-coding genes harbor *cis*-regulatory variants in the same samples, at an FDR of only 0.01 (Barreiro et al. 2012), supports the notion that miRNA expression is under less genetic control than that of mRNAs (Su et al. 2011; Civelek et al. 2013; Lappalainen et al. 2013). The fewer eQTLs detected for miRNAs may reflect a lower permissibility of large changes in their expression, because these would have extensive consequences on the multiple genes and pathways with which miRNAs are associated. Moreover, given the strong conservation observed in miRNA sequences and expression patterns (Quach et al. 2009; Christodoulou et al. 2010; Berezikov 2011), one may also anticipate greater sequence conservation in their regulatory regions, and hence a lower dependence of miRNA expression on proximate genetic variants.

Although several studies have mapped gene-environment interactions for expression phenotypes of protein-coding genes (Smirnov et al. 2009; Romanoski et al. 2010; Maranville et al. 2011; Barreiro et al. 2012; Idaghdour et al. 2012), there has been no effort to characterize the genetic architecture of miRNA expression upon perturbation by external stimuli. Here, we provide evidence of genotype-by-infection interactions that affect miRNA expression variation. The strongest signal among response miR-eQTLs is that observed for miR-326—associated with an eQTL exclusively upon infection (Fig. 2A).

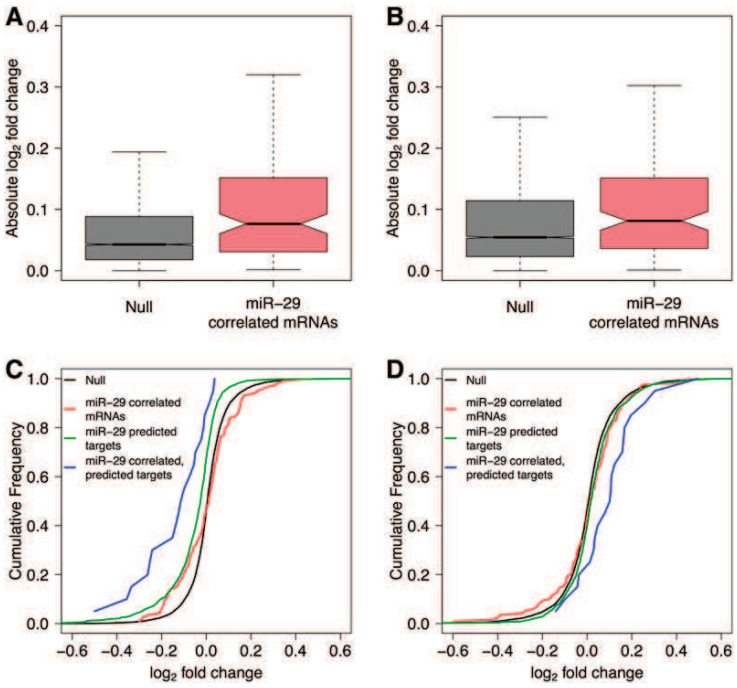



**Figure 4.** Functional validation of miRNA–mRNA relationships using miR-29a gain- and loss-of-function experiments. (*A,B*) Boxplots showing absolute fold changes in genome-wide mRNA levels of cells transfected with either an miR-29a mimic (*A*) or an miR-29 inhibitor (*B*) and subsequently infected with MTB. (*C,D*) Cumulative distributions of changes in genome-wide mRNA levels after transfection with the mimic (*C*) or the inhibitor (*D*) in MTB-infected cells.

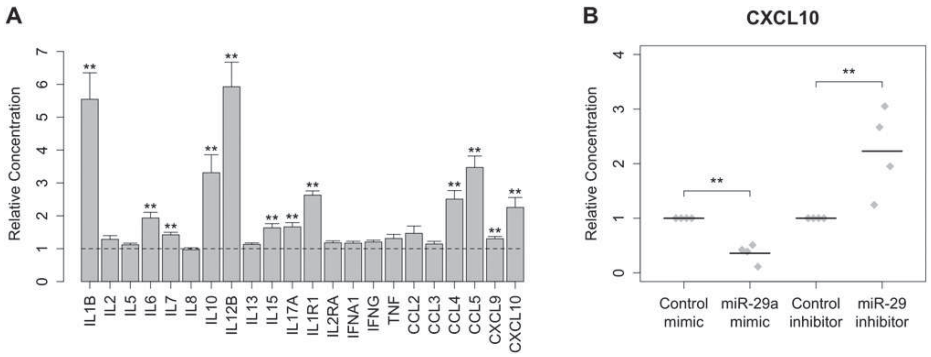



**Figure 5.** Inhibition of miR-29 up-regulates the secretion of multiple cytokines in MTB-infected DCs. Cytokine and chemokine concentrations in culture supernatants from transfected, infected cells were analyzed by multiplex bead immunoassay. (*A*) Relative cytokine/chemokine concentrations in miR-29 inhibitor-transfected cells compared with control-transfected samples ±SEM for duplicates of infection performed on four different donors. MiR-29 inhibition significantly increased the expression of 12 out of 22 cytokines. (**) $P < 0.01$. (*B*) Relative concentrations of the chemokine CXCL10 for mimic- and inhibitor-transfected cells, with respect to control-transfected samples, for duplicates of infection performed on four different donors.

The increased expression of miR-326 in T cells promotes the generation of Th17 cells and is associated with the severity of autoimmune disease (Du et al. 2009). In MTB infection, a shift toward a stronger induction of the Th17 pathway has been associated with excessive neutrophil recruitment, tissue damage, and increased disease severity (Torrado and Cooper 2010; Jurado et al. 2012). The down-regulation of miR-326 that we observed upon infection points to a direct link between this miRNA and DC responses to MTB infection. Furthermore, that a subset of individuals characterized by the *AA* genotype show higher miR-326 expression after infection (~10% of Europeans) (Fig. 2A), which may decrease protective immunity, suggests that regulatory variation at this locus might ultimately impact inter-individual differences in the host response to MTB. This example, together with the infection-dependent eQTL for miR-1260 expression and the interaction effect associated with miR-338-3p, provides experimentally testable hypotheses concerning the role of these miRNAs in immune responses and TB pathogenesis.

The integration of genome-wide expression data from miRNAs and mRNAs highlights the complexity of the miRNA-mediated regulatory system. Although differences in miRNA–mRNA relationships have been reported in diseased and healthy individuals or distinct cell subsets (Cheng et al. 2009; Nunez-Iglesias et al. 2010; Allantaz et al. 2012; Zhang et al. 2012), a systematic study of the impact of an external stimulus on these relationships has been lacking. Our findings showed that infection is accompanied by a rapid and strong remodeling of miRNA-mediated regulatory networks, with a shift toward negative miRNA–mRNA correlations. Such a marked shift, largely accounted for by a small number of differentially expressed miRNAs, emphasizes the wide-reaching impact of a subset of miRNAs in the transcriptional response of a cell to infection.

Through our gain- and loss-of-function experiments, using the model of miR-29a, we have confirmed that the miR-29a-correlated mRNAs detected by our computational analysis are associated with significantly greater changes in expression levels upon perturbation of miR-29 expression in infected cells. This points to a general causal regulation, direct or indirect, of correlated genes by miR-29. That not all genes showed such a change, however, is consistent with the expectation that miRNA–mRNA interactions reflect the coregulation of miRNA–mRNA pairs and/or miRNA sensing as well as signaling (Su et al. 2011). In addition, positively correlated miR-29a predicted targets displayed changes in their expression levels that are consistent with canonical miRNA-mediated repression, lending experimental support to the importance of regulatory loops in miRNA–mRNA interactions (Tsang et al. 2007; Martinez et al. 2008; Ebert and Sharp 2012). Given the large number of positive correlations that we and others report (Martinez et al. 2008; Nunez-Iglesias et al. 2010; Su et al. 2011; Lappalainen et al. 2013), and the enrichment in miRNA predicted targets observed among them, it thus seems likely that feedforward and feedback loops are widespread mechanisms in miRNA-mediated regulatory responses in the context of infection.

Our functional study of miR-29a has also provided new insight into the role of this miRNA in DC functions and responses to MTB infection. First, we observed more differentially expressed genes between miR-29- and control-transfected cells in non-infected conditions, with respect to infected samples. This suggests that, upon infection, miR-29 may drive more focused changes in a smaller set of genes. Second, not only was miR-29a strongly up-regulated in our setting but, most importantly, it had a substantial impact on cytokine secretion in response to infection. In particular, the secretion of CXCL10 is consistent with a direct, repressive effect of miR-29 on this chemokine, which may impact the recruitment of Th1 cells upon MTB infection (Giacomini et al. 2006).

In conclusion, our study has provided an initial assessment of the impact of genotype–environment interactions on miRNA molecular phenotypes by identifying response miR-eQTLs related to infection. This, together with the observed infection-dependent shift of miRNA–mRNA relationships driven by a few miRNAs, such as miR-29a, paves the way for additional studies to evaluate the biological contribution of these miRNAs to immunity to infection and disease outcome.

## Methods

### Samples and miRNA expression analyses

Blood samples were obtained from 65 healthy donors from Research Blood Components. Signed, written consent was obtained from all individuals, in accordance with the company's independent ethics committee approval. Isolation and infection of DCs with *Mycobacterium tuberculosis* (H37Rv) for 18 h, RNA extraction

and quality verification, DNA extraction, and genome-wide genotyping have been previously described (Barreiro et al. 2012). Genome-wide miRNA expression was profiled using the Agilent Human miRNA microarray (Release 16.0). After a series of quality checks, preprocessing, and normalization of the data (Supplemental Methods), we identified differentially expressed miRNAs upon MTB infection by applying a linear model with a fixed effect for MTB treatment, implemented in the Bioconductor package limma (Smyth 2004).

### Mapping of expression quantitative trait loci (eQTLs)

Associations between SNP genotypes (GEO accession number GSE34588) (Barreiro et al. 2012) and miRNA expression levels were calculated using a linear regression model, assuming an additive effect of alleles on expression, in infected and non-infected samples. We improved the power to detect eQTLs by quantile normalization and regression of a number of principal components, to account for unknown confounders (Supplemental Fig. S9). FDRs were estimated by comparing the observed to a null distribution, generated using the lowest *P*-values observed for each miRNA in 100 permutations of expression values (Pickrell et al. 2010; Barreiro et al. 2012). Genotype-treatment interaction effects were detected by Bayesian regression with the software BRIdGE (Maranville et al. 2011). For *trans*-eQTLs, associations were calculated with both genome-wide genotyping data from the same individuals (Barreiro et al. 2012) and a subset of SNPs previously identified as susceptibility loci for TB by GWAs (http://www.genome.gov/26525384/) (Hindorff et al. 2009). Multiple testing corrections were performed using a Bonferroni correction at the 95% significance level. The overlap of miR-eQTLs with active genomic regions was assessed using data from the ENCODE Project (http://encodeproject.org/ENCODE/) (The ENCODE Project Consortium 2012). Enrichments were calculated using a Fisher's exact test. For the fine-mapping of miR-eQTL regions, we imputed genotypes for SNPs not present on our genotyping array with IMPUTE2 (Howie et al. 2009), using integrated haplotype data from Phase 1 of the 1000 Genomes Project (The 1000 Genomes Project Consortium 2012). For details, see the Supplemental Methods.

### Correlation of miRNA and mRNA expression levels

We calculated Pearson correlation coefficients between quantile normalized expression levels of miRNAs and mRNAs detected in at least 50% of samples in at least one condition. As clustered miRNAs show correlated expression profiles (Supplemental Fig. S10), we considered only the most abundant member of each pre-miRNA ($N = 221$). Expression levels of 12,958 protein-coding genes were previously obtained from the same 63 individuals (Barreiro et al. 2012). Significant miRNA–mRNA correlations were determined at an FDR <0.005 based on a null distribution of 1000 permutations and differences between the means of real and null distributions were determined using a *t*-test. Predicted targets of miRNAs were obtained from TargetScan (v6.2) (Friedman et al. 2009). The enrichment of predicted targets of all miRNAs in the network among all miRNA-correlated genes was calculated using a Fisher's exact test. Enrichments of functional Gene Ontology categories and KEGG pathways among the same gene sets were computed using GeneTrail (Backes et al. 2007). We used all 12,958 expressed genes as a background set for over/under-representation analyses of correlated gene sets. Enrichment *P*-values were calculated using a hypergeometric test and we used the Benjamini and Hochberg (1995) approach to correct for multiple testing. Networks were visualized using Cytoscape (Cline et al. 2007).

### Functional analyses of miR-29a using gain- and loss-of-function approaches

Immature DCs from four unrelated individuals were transfected on day 5 using HiPerFect transfection reagent. MiRCURY LNA Power Inhibitors were purchased from Exiqon (miR-29 family 460039, control 199020-00) and miRIDIAN microRNA mimics from Thermo Fisher (miR-29a C-300504-07, control CN-001000-01). Transfection efficiency was assessed by flow cytometry using a fluorescently labeled control oligonucleotide (Exiqon, 199020-04), and found to be on average 77% (Supplemental Fig. S11). Transfected cells were then infected for 24 h with MTB (H37Rv) (Supplemental Fig. S12). MiR-29 expression upon MTB infection and the extent of miR-29 perturbation in transfected cells were quantified by quantitative real-time PCR (qPCR). For gene expression analysis, genome-wide profiling of non-infected and MTB-infected samples was obtained by hybridizing RNA to Illumina HumanHT-12 v4 Expression BeadChip arrays. After a series of quality checks and preprocessing steps, differential expression analysis was performed using the Bioconductor package limma (Smyth 2004). For quantification of cytokine and chemokine levels, we measured supernatant levels of 25 cytokines/chemokines using the Human Cytokine Magnetic 25-Plex Panel (Invitrogen). Differences in secretion levels between conditions were calculated using a Wilcoxon paired rank sum test. For details, see the Supplemental Methods.

## Data access

The miRNA and mRNA expression data reported in this manuscript have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession numbers GSE49951 and GSE53143, respectively.

## Acknowledgments

We thank Robin Friedman for critical reading of the manuscript and useful suggestions; Yoav Gilad for facilitating access to mRNA expression data; and Maud Fagny, Eddie Loh, and John Marioni for helpful discussions. The laboratory of L.Q.-M. has received funding from the Institut Pasteur, the Centre Nationale de la Recherche Scientifique (CNRS), the French government's Investissement d'Avenir program, Laboratoire d'Excellence "Integrative Biology of Emerging Infectious Diseases" (grant no. ANR-10-LABX-62-IBEID), and the European Research Council under the European Union's Seventh Framework Programme (FP/2007–2013)/ERC grant agreement no. 281297. The laboratory of L.B.B. has received funding from the Canadian Institute of Health Research (grant nos. 232519 and 285969) and the Human Frontiers Science Program (grant no. CDA00025-2012). L.B.B. is a scholar of the Fonds de la Recherche en Santé du Québec. The laboratory of L.T. and B.G. has received funding from the National Institutes of Health (grant no. NIH AI087658) and the European Commission within the 7th Framework Programme (grant no. HEALTH-F3-2009-241745). K.J.S. is a scholar of the Pasteur–Paris University (PPU) International PhD program and was supported by a stipend from the Direction Générale de l'Armement (DGA). Y.N. was supported by a fellowship from the Réseau de médecine génétique appliquée, Fonds de recherche du Québec–Santé (FRQS), and the Instituts de recherche en santé du Canada (IRSC, subvention no. TGF-96109).

*Author contributions*: The study was designed by K.J.S., L.B.B., and L.Q.-M. Computational analyses were performed by K.J.S. with contributions and input from Y.N., E.P., G.L., L.B.B., and L.Q.-M. Functional inhibition and overexpression experiments

were performed by M.D. and differentiation and infection of dendritic cells by L.T., with contributions and guidance from J.P., G.L.-V., V.L., and O.N. B.G. contributed with materials and tools. The manuscript was written by K.J.S. and L.Q.-M. with input from all authors.

## References


The 1000 Genomes Project Consortium. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491:** 56–65.

Allantaz F, Cheng DT, Bergauer T, Ravindran P, Rossier MF, Ebeling M, Badi L, Reis B, Bitter H, D'Asaro M, et al. 2012. Expression profiling of human immune cell subsets identifies miRNA-mRNA regulatory relationships correlated with cell type specific expression. *PLoS ONE* **7:** e29979.

Ambros V. 2004. The functions of animal microRNAs. *Nature* **431:** 350–355.

Amit I, Garber M, Chevrier N, Leite AP, Donner Y, Eisenhaure T, Guttman M, Grenier JK, Li W, Zuk O, et al. 2009. Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. *Science* **326:** 257–263.

Backes C, Keller A, Kuentzer J, Kneissl B, Comtesse N, Elnakady YA, Muller R, Meese E, Lenhof HP. 2007. GeneTrail—advanced gene set enrichment analysis. *Nucleic Acids Res* **35:** W186-W192.

Baek D, Villen J, Shin C, Camargo FD, Gygi SP, Bartel DP. 2008. The impact of microRNAs on protein output. *Nature* **455:** 64–71.

Barreiro LB, Tailleux L, Pai AA, Gicquel B, Marioni JC, Gilad Y. 2012. Deciphering the genetic architecture of variation in the immune response to *Mycobacterium tuberculosis* infection. *Proc Natl Acad Sci* **109:** 1204–1209.

Bartel DP. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116:** 281–297.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol* **57:** 289–300.

Berezikov E. 2011. Evolution of microRNA diversity and regulation in animals. *Nat Rev Genet* **12:** 846–860.

Borel C, Deutsch S, Letourneau A, Migliavacca E, Montgomery SB, Dimas AS, Vejnar CE, Attar H, Gagnebin M, Gehrig C, et al. 2011. Identification of *cis*- and *trans*-regulatory variation modulating microRNA expression levels in human fibroblasts. *Genome Res* **21:** 68–73.

Brain O, Owens BM, Pichulik T, Allan P, Khatamzas E, Leslie A, Steevels T, Sharma S, Mayer A, Catuneanu AM, et al. 2013. The intracellular sensor NOD2 induces microRNA-29 expression in human dendritic cells to limit IL-23 release. *Immunity* **39:** 521–536.

Ceppi M, Pereira PM, Dunand-Sauthier I, Barras E, Reith W, Santos MA, Pierre P. 2009. MicroRNA-155 modulates the interleukin-1 signaling pathway in activated human monocyte-derived dendritic cells. *Proc Natl Acad Sci* **106:** 2735–2740.

Chen K, Rajewsky N. 2006. Natural selection on human microRNA binding sites inferred from SNP data. *Nat Genet* **38:** 1452–1456.

Chen CZ, Li L, Lodish HF, Bartel DP. 2004. MicroRNAs modulate hematopoietic lineage differentiation. *Science* **303:** 83–86.

Cheng C, Fu X, Alves P, Gerstein M. 2009. mRNA expression profiles show differential regulatory effects of microRNAs between estrogen receptor-positive and estrogen receptor-negative breast cancer. *Genome Biol* **10:** R90.

Chevrier N, Mertins P, Artyomov MN, Shalek AK, Iannacone M, Ciaccio MF, Gat-Viks I, Tonti E, DeGrace MM, Clauser KR, et al. 2011. Systematic discovery of TLR signaling components delineates viral-sensing circuits. *Cell* **147:** 853–867.

Christodoulou F, Raible F, Tomer R, Simakov O, Trachana K, Klaus S, Snyman H, Hannon GJ, Bork P, Arendt D. 2010. Ancient animal microRNAs and the evolution of tissue identity. *Nature* **463:** 1084–1088.

Civelek M, Hagopian R, Pan C, Che N, Yang WP, Kayne PS, Saleem NK, Cederberg H, Kuusisto J, Gargalovic PS, et al. 2013. Genetic regulation of human adipose microRNA expression and its consequences for metabolic traits. *Hum Mol Genet* **22:** 3023–3037.

Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B, et al. 2007. Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc* **2:** 2366–2382.

Cullen BR. 2011. Viruses and microRNAs: RISCy interactions with serious consequences. *Genes Dev* **25:** 1881–1894.

Degner JF, Pai AA, Pique-Regi R, Veyrieras JB, Gaffney DJ, Pickrell JK, De Leon S, Michelini K, Lewellen N, Crawford GE, et al. 2012. DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature* **482:** 390–394.

Dombroski BA, Nayak RR, Ewens KG, Ankener W, Cheung VG, Spielman RS. 2010. Gene expression and genetic variation in response to endoplasmic reticulum stress in human cells. *Am J Hum Genet* **86:** 719–729.

Du C, Liu C, Kang J, Zhao G, Ye Z, Huang S, Li Z, Wu Z, Pei G. 2009. MicroRNA miR-326 regulates TH-17 differentiation and is associated with the pathogenesis of multiple sclerosis. *Nat Immunol* **10:** 1252–1259.

Ebert MS, Sharp PA. 2012. Roles for microRNAs in conferring robustness to biological processes. *Cell* **149:** 515–524.

The ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489:** 57–74.

Eulalio A, Schulte L, Vogel J. 2012. The mammalian microRNA response to bacterial infections. *RNA Biol* **9:** 742–750.

Friedman RC, Farh KK, Burge CB, Bartel DP. 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* **19:** 92–105.

Fu Y, Yi Z, Wu X, Li J, Xu F. 2011. Circulating microRNAs in patients with active pulmonary tuberculosis. *J Clin Microbiol* **49:** 4246–4251.

Gamazon ER, Ziliak D, Im HK, LaCroix B, Park DS, Cox NJ, Huang RS. 2012. Genetic architecture of microRNA expression: implications for the transcriptome and complex traits. *Am J Hum Genet* **90:** 1046–1063.

Gamazon ER, Innocenti F, Wei R, Wang L, Zhang M, Mirkov S, Ramirez J, Huang RS, Cox NJ, Ratain MJ, et al. 2013. A genome-wide integrative study of microRNAs in human liver. *BMC Genomics* **14:** 395.

Gargalovic PS, Imura M, Zhang B, Gharavi NM, Clark MJ, Pagnon J, Yang WP, He A, Truong A, Patel S, et al. 2006. Identification of inflammatory gene modules based on variations of human endothelial cell responses to oxidized lipids. *Proc Natl Acad Sci* **103:** 12741–12746.

Gat-Viks I, Chevrier N, Wilentzik R, Eisenhaure T, Raychowdhury R, Steuerman Y, Shalek AK, Hacohen N, Amit I, Regev A. 2013. Deciphering molecular circuits from genetic variation underlying transcriptional responsiveness to stimuli. *Nat Biotechnol* **31:** 342–349.

Giacomini E, Iona E, Ferroni L, Miettinen M, Fattorini L, Orefici G, Julkunen I, Coccia EM. 2001. Infection of human macrophages and dendritic cells with *Mycobacterium tuberculosis* induces a differential cytokine gene expression that modulates T cell response. *J Immunol* **166:** 7033–7041.

Giacomini E, Sotolongo A, Iona E, Severa M, Remoli ME, Gafa V, Lande R, Fattorini L, Smith I, Manganelli R, et al. 2006. Infection of human dendritic cells with a *Mycobacterium tuberculosis* sigE mutant stimulates production of high levels of interleukin-10 but low levels of CXCL10: impact on the T-cell response. *Infect Immun* **74:** 3296–3304.

Gilad Y, Rifkin SA, Pritchard JK. 2008. Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends Genet* **24:** 408–415.

Guo H, Ingolia NT, Weissman JS, Bartel DP. 2010. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466:** 835–840.

Hickman SP, Chan J, Salgame P. 2002. *Mycobacterium tuberculosis* induces differential cytokine production from dendritic cells and macrophages with divergent effects on naive T cell polarization. *J Immunol* **168:** 4636–4642.

Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci* **106:** 9362–9367.

Hornstein E, Shomron N. 2006. Canalization of development by microRNAs. *Nat Genet* (Suppl) **38:** S20–S24.

Howie BN, Donnelly P, Marchini J. 2009. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5:** e1000529.

Huang Q, Liu D, Majewski P, Schulte LC, Korn JM, Young RA, Lander ES, Hacohen N. 2001. The plasticity of dendritic cell responses to pathogens and their components. *Science* **294:** 870–875.

Huang RS, Gamazon ER, Ziliak D, Wen Y, Im HK, Zhang W, Wing C, Duan S, Bleibel WK, Cox NJ, et al. 2011. Population differences in microRNA expression and biological implications. *RNA Biol* **8:** 692–701.

Huntzinger E, Izaurralde E. 2011. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat Rev Genet* **12:** 99–110.

Idaghdour Y, Quinlan J, Goulet JP, Berghout J, Gbeha E, Bruat V, de Malliard T, Grenier JC, Gomez S, Gros P, et al. 2012. Evidence for additive and interaction effects of host genotype and infection in malaria. *Proc Natl Acad Sci* **109:** 16786–16793.

Johnnidis JB, Harris MH, Wheeler RT, Stehling-Sun S, Lam MH, Kirak O, Brummelkamp TR, Fleming MD, Camargo FD. 2008. Regulation of progenitor cell proliferation and granulocyte function by microRNA-223. *Nature* **451:** 1125–1129.

Jurado JO, Pasquinelli V, Alvarez IB, Pena D, Rovetta AI, Tateosian NL, Romeo HE, Musella RM, Palmero D, Chuluyan HE, et al. 2012. IL-17 and IFN-γ expression in lymphocytes from patients with active tuberculosis correlates with the severity of the disease. *J Leukoc Biol* **91:** 991–1002.

Krek A, Grun D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da Piedade I, Gunsalus KC, Stoffel M, et al. 2005. Combinatorial microRNA target predictions. *Nat Genet* **37:** 495–500.

Lappalainen T, Sammeth M, Friedlander MR, 't Hoen PA, Monlong J, Rivas MA, Gonzalez-Porta M, Kurbatova N, Griebel T, Ferreira PG, et al. 2013. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501:** 506–511.
Liu G, Friggeri A, Yang Y, Park YJ, Tsuruta Y, Abraham E. 2009. miR-147, a microRNA that is induced upon Toll-like receptor stimulation, regulates murine macrophage inflammatory responses. *Proc Natl Acad Sci* **106:** 15819–15824.
Liu X, Zhan Z, Xu L, Ma F, Li D, Guo Z, Li N, Cao X. 2010. MicroRNA-148/152 impair innate response and antigen presentation of TLR-triggered dendritic cells by targeting CaMKIIα. *J Immunol* **185:** 7244–7251.
Lodish HF, Zhou B, Liu G, Chen CZ. 2008. Micromanagement of the immune system by microRNAs. *Nat Rev Immunol* **8:** 120–130.
Lu J, Clark AG. 2012. Impact of microRNA regulation on variation in human gene expression. *Genome Res* **22:** 1243–1254.
Ma F, Xu S, Liu X, Zhang Q, Xu X, Liu M, Hua M, Li N, Yao H, Cao X. 2011. The microRNA miR-29 controls innate and adaptive immune responses to intracellular bacterial infection by targeting interferon-γ. *Nat Immunol* **12:** 861–869.
Maertzdorf J, Weiner J III, Mollenkopf HJ, Bauer T, Prasse A, Muller-Quernheim J, Kaufmann SH. 2012. Common patterns and disease-related signatures in tuberculosis and sarcoidosis. *Proc Natl Acad Sci* **109:** 7853–7858.
Majewski J, Pastinen T. 2011. The study of eQTL variations by RNA-seq: from SNPs to phenotypes. *Trends Genet* **27:** 72–79.
Maranville JC, Luca F, Richards AL, Wen X, Witonsky DB, Baxter S, Stephens M, Di Rienzo A. 2011. Interactions between glucocorticoid treatment and *cis*-regulatory polymorphisms contribute to cellular response phenotypes. *PLoS Genet* **7:** e1002162.
Martinez NJ, Ow MC, Barrasa MI, Hammell M, Sequerra R, Doucette-Stamm L, Roth FP, Ambros VR, Walhout AJ. 2008. A *C. elegans* genome-scale microRNA network contains composite feedback motifs with high flux capacity. *Genes Dev* **22:** 2535–2549.
Montgomery SB, Dermitzakis ET. 2011. From expression QTLs to personalized transcriptomics. *Nat Rev Genet* **12:** 277–282.
Nica AC, Montgomery SB, Dimas AS, Stranger BE, Beazley C, Barroso I, Dermitzakis ET. 2010. Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet* **6:** e1000895.
Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. 2010. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet* **6:** e1000888.
Nunez-Iglesias J, Liu CC, Morgan TE, Finch CE, Zhou XJ. 2010. Joint genome-wide profiling of miRNA and mRNA expression in Alzheimer's disease cortex reveals altered miRNA regulation. *PLoS ONE* **5:** e8898.
O'Connell RM, Rao DS, Baltimore D. 2012. microRNA regulation of inflammatory responses. *Annu Rev Immunol* **30:** 295–312.
Okamura K, Hagen JW, Duan H, Tyler DM, Lai EC. 2007. The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila*. *Cell* **130:** 89–100.
O'Neill LA, Sheedy FJ, McCoy CE. 2011. MicroRNAs: the fine-tuners of Toll-like receptor signalling. *Nat Rev Immunol* **11:** 163–175.
Parts L, Hedman AK, Keildson S, Knights AJ, Abreu-Goodger C, van de Bunt M, Guerra-Assuncao JA, Bartonicek N, van Dongen S, Magi R, et al. 2012. Extent, causes, and consequences of small RNA expression variation in human adipose tissue. *PLoS Genet* **8:** e1002704.
Pickrell JK, Marioni JC, Pai AA, Degner JF, Engelhardt BE, Nkadori E, Veyrieras JB, Stephens M, Gilad Y, Pritchard JK. 2010. Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* **464:** 768–772.
Qi Y, Cui L, Ge Y, Shi Z, Zhao K, Guo X, Yang D, Yu H, Cui L, Shan Y, et al. 2012. Altered serum microRNAs as biomarkers for the early diagnosis of pulmonary tuberculosis infection. *BMC Infect Dis* **12:** 384.
Quach H, Barreiro LB, Laval G, Zidane N, Patin E, Kidd KK, Kidd JR, Bouchier C, Veuille M, Antoniewski C, et al. 2009. Signatures of purifying and local positive selection in human miRNAs. *Am J Hum Genet* **84:** 316–327.
Rajaram MV, Ni B, Morris JD, Brooks MN, Carlson TK, Bakthavachalu B, Schoenberg DR, Torrelles JB, Schlesinger LS. 2011. *Mycobacterium tuberculosis* lipomannan blocks TNF biosynthesis by regulating macrophage MAPK-activated protein kinase 2 (MK2) and microRNA miR-125b. *Proc Natl Acad Sci* **108:** 17408–17413.
Rantalainen M, Herrera BM, Nicholson G, Bowden R, Wills QF, Min JL, Neville MJ, Barrett A, Allen M, Rayner NW, et al. 2011. MicroRNA expression in abdominal and gluteal adipose tissue is associated with mRNA expression levels and partly genetically driven. *PLoS ONE* **6:** e27338.
Romanoski CE, Lee S, Kim MJ, Ingram-Drake L, Plaisier CL, Yordanova R, Tilford C, Guan B, He A, Gargalovic PS, et al. 2010. Systems genetics analysis of gene-by-environment interactions in human cells. *Am J Hum Genet* **86:** 399–410.
Ruby JG, Jan CH, Bartel DP. 2007. Intronic microRNA precursors that bypass Drosha processing. *Nature* **448:** 83–86.
Saunders MA, Liang H, Li WH. 2007. Human polymorphism at microRNAs and microRNA target sites. *Proc Natl Acad Sci* **104:** 3300–3305.
Selbach M, Schwanhausser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N. 2008. Widespread changes in protein synthesis induced by microRNAs. *Nature* **455:** 58–63.
Sharbati J, Lewin A, Kutz-Lohroff B, Kamal E, Einspanier R, Sharbati S. 2011. Integrated microRNA-mRNA-analysis of human monocyte derived macrophages upon *Mycobacterium avium* subsp. *hominissuis* infection. *PLoS ONE* **6:** e20258.
Smirnov DA, Morley M, Shin E, Spielman RS, Cheung VG. 2009. Genetic analysis of radiation-induced changes in human gene expression. *Nature* **459:** 587–591.
Smith EN, Kruglyak L. 2008. Gene-environment interaction in yeast gene expression. *PLoS Biol* **6:** e83.
Smyth GK. 2004. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3:** Article3.
Stark A, Brennecke J, Bushati N, Russell RB, Cohen SM. 2005. Animal microRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. *Cell* **123:** 1133–1146.
Su WL, Kleinhanz RR, Schadt EE. 2011. Characterizing the role of miRNAs within gene regulatory networks using integrative genomics techniques. *Mol Syst Biol* **7:** 490.
Thye T, Vannberg FO, Wong SH, Owusu-Dabo E, Osei I, Gyapong J, Sirugo G, Sisay-Joof F, Enimil A, Chinbuah MA, et al. 2010. Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. *Nat Genet* **42:** 739–741.
Thye T, Owusu-Dabo E, Vannberg FO, van Crevel R, Curtis J, Sahiratmadja E, Balabanova Y, Ehmen C, Muntau B, Ruge G, et al. 2012. Common variants at 11p13 are associated with susceptibility to tuberculosis. *Nat Genet* **44:** 257–259.
Torrado E, Cooper AM. 2010. IL-17 and Th17 cells in tuberculosis. *Cytokine Growth Factor Rev* **21:** 455–462.
Tsang J, Zhu J, van Oudenaarden A. 2007. MicroRNA-mediated feedback and feedforward loops are recurrent network motifs in mammals. *Mol Cell* **26:** 753–767.
Tsang JS, Ebert MS, van Oudenaarden A. 2010. Genome-wide dissection of microRNA functions and cotargeting networks using gene set signatures. *Mol Cell* **38:** 140–153.
Turner ML, Schnorfeil FM, Brocker T. 2011. MicroRNAs regulate dendritic cell differentiation and function. *J Immunol* **187:** 3911–3917.
Vasudevan S, Tong Y, Steitz JA. 2007. Switching from repression to activation: microRNAs can up-regulate translation. *Science* **318:** 1931–1934.
Wang L, Oberg AL, Asmann YW, Sicotte H, McDonnell SK, Riska SM, Liu W, Steer CJ, Subramanian S, Cunningham JM, et al. 2009. Genome-wide transcriptional profiling reveals microRNA-correlated genes and biological processes in human lymphoblastoid cell lines. *PLoS ONE* **4:** e5878.
Xia K, Shabalin AA, Huang S, Madar V, Zhou YH, Wang W, Zou F, Sun W, Sullivan PF, Wright FA. 2012. seeQTL: a searchable database for human eQTLs. *Bioinformatics* **28:** 451–452.
Yang IV, Wade CM, Kang HM, Alper S, Rutledge H, Lackford B, Eskin E, Daly MJ, Schwartz DA. 2009. Identification of novel genes that mediate innate immunity using inbred mice. *Genetics* **183:** 1535–1544.
Yi Z, Fu Y, Ji R, Li R, Guan Z. 2012. Altered microRNA signatures in sputum of patients with active pulmonary tuberculosis. *PLoS ONE* **7:** e43184.
Zhang W, Edwards A, Fan W, Flemington EK, Zhang K. 2012. miRNA-mRNA correlation-network modules in human prostate cancer and the differences between primary and metastatic tumor subtypes. *PLoS ONE* **7:** e40130.

*Received June 3, 2013; accepted in revised form January 22, 2014.*

## 2.3. Résumé des travaux et discussion spécifique

La susceptibilité aux maladies infectieuses est un phénotype excessivement complexe, variable entre individus et dont la plupart des mécanismes moléculaires sous-jacents restent mal décrits. Dans cette étude, nous avons pour la première fois identifié des différences de réponses impliquant les miARN dans les cellules dendritiques humaines suite à l'infection par *Mycobacterium tuberculosis*, quantifié l'impact de ces réponses sur la régulation de l'expression des gènes codant pour des protéines et caractérisé certains facteurs génétiques liés à cette variabilité d'expression. Ainsi, nous avons tout d'abord montré que l'expression d'environ 40 % des miARN était modifiée dans les cellules dendritiques suite à l'infection par la tuberculose. Nous avons pu retrouver de nombreux résultats précédemment décrits, confirmant ainsi notre approche, tout en mettant pour la première fois en évidence l'implication de certains miARN dans la réponse des cellules dendritiques à l'infection par la tuberculose.

Par la suite, nous avons cherché à identifier des variations génétiques influençant l'expression des miARN dans au moins une condition et à caractériser les régions génomiques dans lesquelles elles se trouvent. Nous avons montré que les miR-eQTL identifiés sont enrichis en régions associées à des marques épigénétiques caractéristiques d'une activation de la transcription, suggérant que les variations génétiques que nous décrivons sont d'une importance majeure pour l'expression des miARN.

Finalement, nous avons estimé la relation entre miARN et ARNm avant et après infection des cellules dendritiques par MTB. Nous avons remarqué que l'expression des miARN était corrélée à un plus grand nombre de transcrits suite à l'infection et que cette corrélation était principalement due à un faible nombre de miARN, nous permettant d'établir les principales connexions du réseau d'interaction miARN-ARNm et d'en estimer la plasticité au cours de la stimulation des cellules par un agent infecieux. Nous avons validé certaines de ces observations à l'aide d'études de gain ou perte de fonction en utilisant miR-29 comme modèle. Cette dernière approche nous a aussi permis de caractériser de nouvelles fonctions de miR-29 dans la réponse transcriptionnelle et protéique des cellules dendritiques à l'infection par MTB.

Notre approche génome entier a permis d'identifier les dérégulations de l'expression de miARN qui n'avaient jusqu'à présent pas été décrites dans la réponse immunitaire. Notre approche suggère donc que ces transcrits sont importants pour la réponse des cellules

dendritiques à l'infection. De plus amples études permettraient d'identifier leur fonction moléculaire dans ce contexte et, plus largement, dans l'immunité.

Nous avons ensuite montré que les corrélations entre variations génétiques et niveaux d'expression des miARN étaient moins nombreuses que celles impliquant les ARNm. Les miARN sont très conservés (Quach et al., 2009), régulent l'expression de nombreuses cibles (REF) et sont impliqués dans un large spectre de processus cellulaires. Nous supposons donc que le nombre restreint de miR-eQTL reflète une faible permissivité de variation de l'expression des miARN due aux larges conséquences que ces modifications auraient sur un grand nombre de cibles et, plus généralement, sur un ensemble de voies de signalisation. Selon cette hypothèse, il apparaît d'importance majeure de caractériser le rôle des miARN dont nous décrivons l'abondance comme étant sous contrôle génétique et, plus particulièrement, des transcrits dont l'expression dépend d'une combinaison de facteurs génétiques et environnementaux. Les rôles précédemment décrits de miR-326, l'un des candidats remplissant ces critères, dans d'autres cellules immunitaires suggèrent que ce miARN pourrait être la clé d'un des mécanismes de différence de susceptibilité aux maladies infectieuses.

Nous avons confirmé l'étendue du remodelage des interactions miARN-ARNm au cours de l'infection et la complexité du réseau de régulation de l'expression génique par l'intégration de l'expression de ces deux types de transcrits ainsi que par des analyses moléculaires se concentrant sur miR-29. Cette dernière étude a non seulement été utilisée comme validation expérimentale de notre analyse pangénomique mais nous a aussi permis de caractériser certaines des fonctions de ce miARN dans la réponse transcriptionnelle et protéique des cellules dendritiques à l'infection par MTB. Nous avons ainsi montré que miR-29 régulait la quantité de CXCL10, une chimiokine particulièrement importante pour le recrutement des lymphocytes T sur le site de l'infection. Cela suggère que la différence d'expression de miR-29 dans les cellules dendritiques pourrait permettre la régulation de mécanismes cellulaires complexes.

Notre étude prise dans son ensemble apporte donc un éclairage nouveau sur le rôle des miARN dans l'immunité et identifie des candidats pouvant influencer la variabilité de réponse immunitaire de l'hôte suite à l'infection par MTB.

# Discussion générale et perspectives

## 1. La reconnaissance des acides nucléiques, mécanisme clé de l'immunité innée ?

Notre étude a permis de confirmer l'étendue de l'action de la sélection purificatrice sur les gènes de l'immunité innée. Cependant, il existe une importante hétérogénéité de l'intensité de ce type de sélection, tant au niveau des gènes pris individuellement qu'au sein même des différentes classes fonctionnelles que nous avons définies.

C'est notamment le cas pour le groupe des récepteurs de l'immunité innée, qui ont pour fonction de reconnaître des motifs moléculaires spécifiques conservés chez les micro-organismes et qui montrent une forte disparité de contrainte évolutive. Il a été précédemment observé que certaines des molécules servant à la reconnaissance de motifs microbiens dans le cytosol sont particulièrement conservées (Barreiro et al., 2009; Vasseur et al., 2012; Wlasiuk and Nachman, 2010). Dans le cadre de la reconnaissance des acides nucléiques, ces protéines reconnaissent des motifs particuliers du matériel génétique viral ou bactérien, comme la structure en double brin des ARN ou les îlots CpG déméthylés. Si ces propriétés physico-chimiques permettent de distinguer les acides nucléiques de pathogènes de ceux de l'hôte, une variation génétique modifiant la structure du senseur peut rapidement mener à une reconnaissance de motifs spécifiques à l'hôte, aboutissant à une réponse auto-immune (Deane and Bolland, 2006; Krieg and Vollmer, 2007; Marshak-Rothstein, 2006). Ces constatations ont mené à l'hypothèse que la limite très fine existant entre la reconnaissance des acides nucléiques des pathogènes et ceux de l'hôte serait à l'origine de la conservation particulièrement importante des TLR localisés au niveau des endosomes.

De manière particulièrement marquante, nous avons montré dans nos travaux que la classe de récepteurs la plus contrainte était celles des senseurs des acides nucléiques cytosoliques. Ainsi, les cinq récepteurs les plus contraints reconnaissent les ARN ou ADN de microbiens: DHX9 fixe les ADN contenant des motifs CpG dans le cytosol et est important pour la réponse impliquant l'interféron dans les cellules dendritiques plasmacytoïdes suite à l'infection par l'*Herpes simplex virus* 1 (Kim et al., 2010); DHX15 reconnaît les ARN double-brins dans les cellules dendritiques myéloïdes (Lu et al., 2014); DDX41 se fixe aux messagers secondaires bactériens et induit une réponse médiée par les interférons de type 1 (Parvatiyar et al., 2012); HMGB2 est un senseur des ADN cytosoliques double-brins (Yanai et al., 2009) et DDX3, outre son implication dans de multiple processus cellulaires, a été proposé comme une protéine reconnaissant les ARN pathogéniques (Schröder et al., 2008).

L'hypothèse qui epliquerait la forte pression de sélection exercée sur les senseurs des acides nucléiques par la nécessité d'une balance entre reconnaissance efficace des pathogènes et reconnaissance du matériel génétique de l'hôte semble donc s'élargir à d'autres groupes que les TLR. Il serait intéressant de procéder à des analyses similaires chez d'autres primates. En effet, l'intégration de données provenant de plusieurs organismes permettrait de déterminer si ces gènes sont universellement contraints ou s'ils ont participé à l'adaptation de notre espèce *via* un évènement de divergence suivi de l'action de la sélection purificatrice.

Cependant, en se basant sur l'exemple de TLR8 qui est particulièrement conservé et qui a été rapporté comme étant impliqué dans la croissance neuronale dans des modèles murins (Ma et al., 2007, 2006), il n'est pas à exclure que les CNAS sont impliqués dans des fonctions cellulaires autres que l'immunité. DDX3 a par exemple été étudié dans de nombreux processus cellulaires tels que le métabolisme des ARNm (Ariumi, 2014; Soto-Rifo and Ohlmann, 2013), qui sont indispensables au bon fonctionnement de la cellule. Une meilleure caractérisation des fonctions moléculaires de cette classe nouvellement définie de récepteurs de l'immunité innée est donc nécessaire. Si ces molécules s'avèrent impliquées dans de multiples processus cellulaires, une estimation du rôle de leur implication dans l'immunité innée dans l'intensité de sélection qui est exercée sur ces gènes serait nécessaire. Plus généralement, ce type d'étude pourrait être appliqué à l'ensemble des gènes de l'immunité innée et permettrait de déterminer si ce processus cellulaire est celui expliquant principalement l'importante conservation des gènes de l'immunité innée.

## 2. Gènes de l'immunité innée et autres modes de sélection positive

Notre analyse a montré que les gènes impliqués dans la réponse immunitaire innée n'ont pas été des cibles privilégiées de la sélection positive. Cependant, il est important de noter que notre étude est principalement axée sur l'identification d'évènements de sélection à fort coefficient touchant des allèles à faibles fréquences dans la population avant le début des évènements de sélection. Afin d'apprécier l'action de la sélection positive sur les gènes de l'immunité innée dans son ensemble, il faudrait considérer d'autres types de sélection positive : la sélection à plus faible coefficient, la sélection sur des variations génétiques préexistantes à fréquences modérées dans la population ainsi que la sélection polygénique (Fu and Akey, 2013; Hermisson and Pennings, 2005; Messer and Petrov, 2013; Pritchard et al., 2010; Pritchard and Di Rienzo, 2010; Scheinfeldt and Tishkoff, 2013).

Ainsi, une mutation peut apparaître *de novo* à de multiples reprises ou être présente dans la population associée à diverses combinaisons alléliques avant d'être sujette à l'action de la sélection positive. Dans ces cas, plusieurs haplotypes avantageux vont avoir une fréquence relativement élevée dans la population étudiée (Hermisson and Pennings, 2005; Pennings and Hermisson, 2006a, 2006b; Przeworski et al., 2005). Les signatures locales de diversité génétique et de DL sont alors différentes de celles obtenues par balayage sélectif, les rendant difficiles à identifier à l'aide des outils classiques (Pennings and Hermisson, 2006b; Pritchard et al., 2010; Przeworski et al., 2005; Teshima et al., 2006). Cependant, ces balayages de faible intensité semblent constituer l'un des mécanismes d'action de la sélection naturelle permettant aux organismes de s'adapter à leur environnement (Messer and Petrov, 2013; Pritchard et al., 2010). L'utilisation de méthodes composites, tout comme celle de tests de différenciation entre populations, permet en théorie d'identifier des évènements de sélection locale agissant sur des variations génétiques déjà présentes dans la population (Grossman et al., 2010; Innan and Kim, 2008). Malgré tout, le développement d'outils permettant de détecter les balayages sélectifs de faible intensité est nécessaire. De récents travaux allant dans ce sens ont récemment été menés en utilisant la drosophile comme organisme modèle (Garud et al., 2015; Garud and Rosenberg, 2015). L'application de telles statistiques chez l'Homme serait donc une première étape vers une meilleure appréciation de l'action de la sélection positive dans son ensemble sur les gènes de l'immunité innée.

La détection de la sélection polygénique est aussi l'un des axes voués à se développer au cours de ces prochaines années. Ce type de sélection, qui consiste en l'accumulation de multiples modifications touchant plusieurs gènes impliqués dans le même processus cellulaire, chacune ayant individuellement peu d'effet sur le phénotype final, pourrait constituer l'un des principaux mécanismes d'adaptation des organismes à leur environnement (Hernandez et al., 2011; Messer and Petrov, 2013). De nombreux traits phénotypiques comme la résistance aux pathogènes ou la taille sont le résultat de l'expression de nombreux gènes, en faisant des cibles privilégiées de l'action de la sélection polygénique (Daub et al., 2013; Lango Allen et al., 2010; Mendizabal et al., 2012). Plusieurs types d'approches ont été utilisés pour mettre en évidence l'action d'un tel type de sélection. Le premier consiste à rechercher un ensemble de SNP associés au même trait phénotypique (Turchin et al., 2012; Zhang et al., 2013). La corrélation entre paramètres environnementaux et fréquences alléliques à différents sites est une autre méthode. Cette approche a notamment permis de montrer que les pathogènes ont constitué la principale force sélective influençant l'évolution

de l'Homme (Fumagalli et al., 2011). Enfin, il est possible de chercher un enrichissement de signaux de sélection positive dans des gènes appartenant à une même voie de signalisation. Une étude a ainsi identifié une accumulation de signaux de sélection positive dans certaines voies de l'immunité telles que la voie de l'IL-6 et des interactions entre les cytokines et leurs récepteurs (Daub et al., 2013). Notre analyse se basant sur une classification fonctionnelle, nous n'avons pu identifier de tels signaux. Reconsidérer notre catégorisation en voies de signalisation pourrait permettre de mettre en évidence d'autres signaux de sélection polygénique.

## 3. Sélection positive et régions régulatrices

Notre étude s'est attachée à identifier des signaux de sélection positive localisés dans des séquences codantes impliquées dans la réponse immunitaire innée. Cependant, nous avons aussi détecté de nombreux signaux dans des régions intergéniques qui sont parfois localisées à proximité de gènes jouant un rôle dans notre processus cellulaire d'intérêt.

Il a été proposé de manière récurrente que les régions régulant l'expression des gènes pourraient avoir joué un rôle majeur dans l'adaptation des populations à leur environnement (ENCODE Project Consortium, 2012; Grossman et al., 2013; King and Wilson, 1975; Wray, 2007). L'exemple de la persistance de l'activité enzymatique de la lactase, que nous avons évoqué dans l'introduction, en est probablement l'exemple le plus parlant. Une récente étude a estimé que les balayages sélectifs touchent majoritairement les régions régulatrices et non les variations génétiques induisant une modification de la séquence d'acides aminés (Fraser, 2013). Les résultats de deux analyses réalisées ces dernières années corroborent cette idée puisqu'une grande partie des régions génomiques identifiées comme étant cibles de la sélection positive correspondent à des régions régulant l'expression des gènes (Enard et al., 2014; Grossman et al., 2013). D'autres travaux ont, eux, trouvé un enrichissement de signaux de sélection positive dans les variations génétiques associées à des différences d'expression géniques (Kudaravalli et al., 2009; Ye et al., 2013). Certaines de ces variations sont notamment impliquées dans des fonctions immunes (Ye et al., 2013), ce qui montre l'importance de considérer les pressions de sélections ayant agi sur les régions régulatrices dans le but de mieux apprécier l'histoire adaptative des populations humaines aux pathogènes présents dans leurs environnements respectifs.

Les efforts importants réalisés ces dernières années ont permis de mieux décrire les propriétés des régions régulant l'expression des gènes et d'affiner leur cartographie. Diverses approches ont été utilisées, aboutissant à l'identification de nombreux promoteurs et enhancers (Belton et al., 2012; ENCODE Project Consortium, 2012; Heintzman et al., 2007; D. S. Johnson et al., 2007; Kim et al., 2005; Niu et al., 2014; Song et al., 2011). Il est donc aujourd'hui possible de déterminer quelles sont les mutations sous sélection positive qui sont localisées dans des régions régulatrices de l'expression des gènes (Grossman et al., 2013). La principale difficulté réside ensuite à identifier la (ou les) variation(s) génétique(s) ayant fonctionnellement participé à l'adaptation de la population d'intérêt à son environnement. Parmi les nombreuses stratégies disponibles (Scheinfeldt and Tishkoff, 2013; Ward and Kellis, 2012) nous avons choisi de croiser nos résultats avec des données d'associations pangénomiques (GWAS). D'autres analyses ont utilisé des données d'expression géniques (Grossman et al., 2013; Kudaravalli et al., 2009) ou des données fonctionnelles comme la sensibilité à la DNase-I, qui est une signature des sites de fixation des protéines participant à la régulation de l'expression des gènes (Vernot et al., 2012). Cependant, la plupart de ces approches *in* silico ne fournissent que des arguments indirects et une validation formelle des conséquences phénotypiques des allèles étudiés doit encore être apportée par des études fonctionnelles.

## 4. Fonctionnalité des allèles archaïques introgressés dans les populations humaines actuelles

Nos travaux montrent qu'une forte introgression d'allèles provenant de génomes archaïques s'est faite dans les gènes de l'immunité innée relativement aux autres régions géniques. Une quantité croissante de données suggère que des gènes impliqués dans de nombreux processus cellulaires dont la réponse immunitaire sont les cibles d'introgression d'allèles provenant d'homininés aujourd'hui éteints et que la diversité génétique ainsi apportée a pu participer à l'adaptation des populations (Abi-Rached et al., 2011; Ding et al., 2014b; Huerta-Sánchez et al., 2014; Khrameeva et al., 2014; Mendez et al., 2012a; Mendez et al., 2012b; Mendez et al., 2013; Nakaoka and Inoue, 2015; Racimo et al., 2015).

Certaines études ont mené à la reconstruction des haplotypes néandertaliens ou dénisoviens qui ont été introgressés chez l'Homme. Nombre de ces combinaisons alléliques portent des allèles modifiant la séquences d'acides aminés des protéines (Abi-Rached et al.,

2011; Huerta-ScitationID":"rpjde0vqv","properties":{"forma qui sont supposées avoir été les cibles de la sélection qui s'est par la suite faite sur ces régions génomiques. Pour autant, l'identification de la (ou des) variation(s) génétique(s) avantageuse(s) héritée(s) des Hommes archaïques reste difficile. En effet, certains des haplotypes ainsi reconstruits couvrent plusieurs régions codantes, ne permettant pas de conclure de manière formelle sur la cible de la sélection positive (l'haplotype néandertalien introgressé dans la région génomique où se situe *STAT2* (Mendez et al., 2012b) englobe aussi *ERBB3* et *ESYT1*). De futurs travaux devront donc viser à améliorer l'identification des allèles introgressés ayant conféré un avantage sélectif à l'Homme moderne.

## 5. eQTL de réponses et stimulations immunitaires

Le deuxième axe développé au cours de cette thèse vise à étudier la variabilité de réponse immunitaire par une approche *ex vivo*. Nos résultats montrent que l'expression d'une partie des miARN dans les cellules dendritiques est corrélée à des variations génétiques localisées à proximité de ces séquences. Une quantité croissante de données accumulées au cours de ces dernières années montre que les effets qu'ont certaines variations génétiques sur l'abondance de transcrits peuvent varier en fonction de variables environnementales ou de traitements (Barreiro et al., 2012; Fairfax et al., 2014; Idaghdour et al., 2012; Lee et al., 2014; Maranville et al., 2011; Romanoski et al., 2010; Smirnov et al., 2009). Notre étude teste pour la première fois cette hypothèse sur des gènes non-codants. Ainsi, nous identifions deux eQTL de réponse dans le cadre de la réponse des celllules dendritiques impliquant les miARN suite à l'infection par *Mycobacterium tuberculosis*. L'ensemble de ces données indique que l'interaction entre gènes et environnement joue un rôle dans la régulation de la quantité de transcrits dans les cellules. Il paraît alors important d'estimer l'impact d'autres facteurs environnementaux sur la régulation de l'expression génique. Pour se restreindre au contexte de l'immunité, deux types d'approches peuvent être utilisés. Le premier consiste à stimuler des cellules immunitaires à l'aide de molécules activant des voies de signalisation spécifiques et, en parallèle, d'un agent infectieux. Ce type d'expériences permet d'estimer la participation de chacune des voies de signalisation étudiées dans la réponse finale à l'infection, d'évaluer les interactions entre ces voies et d'apprécier le degré de variabilité interindividuelle et/ou inter-populationnelles qu'elles présentent. Certaines études ont utilisé des effecteurs produits lors de la réponse immunitaire afin de stimuler des cellules immunitaires (Fairfax et al., 2014; Lee et al., 2014). Cependant, cela ne permet pas d'apprécier les interactions qui peuvent se

faire entre les différents PAMP et les voies PRR. La seconde approche revient à infecter des cellules immunitaires avec un panel plus large de pathogènes (voir section suivante). Ces approches permettraient de proposer de nouveaux sites impliqués dans la variabilité inter-individuelle de réponse immunitaire et de susceptibilité aux maladies infectieuses.

## 6. Étendue et spécificité de la réponse immunitaire impliquant les miARN

Notre étude ainsi que d'autres travaux montrent qu'un grand nombre de miARN est impliqué dans la réponse aux infections par différentes mycobactéries (Fu et al., 2011; Liu et al., 2011; Maertzdorf et al., 2012; Ma et al., 2011; Rajaram et al., 2011; Sharbati et al., 2011; Spinelli et al., 2013; Wu et al., 2012). Une comparaison de ces différentes données pourrait apporter d'importantes informations quant à la contribution des miARN à la réponse immunitaire aux pathogènes. Cela permettrait par exemple de définir un groupe de miARN impliqués dans la régulation du processus immunitaire quel que soit le pathogène considéré ainsi que d'identifier des signatures d'expression des miARN spécifiques à l'infection par une bactérie. Nous montrons par exemple que miR-29 est surexprimé dans les cellules dendritiques suite à l'infection par *Mycobacterium tuberculosis* alors que les résultats obtenus par Ma et al., 2011 attestent d'une diminution de la quantité de ce miARN dans différents types cellulaires suite l'infection de souris par *Mycobacterium bovis* et *Listeria monocytogenes*. Il apparait clairement que les deux analyses ne sont pas comparables. L'organisme modèle, les types cellulaires, les agents infectieux et les temps d'exposition aux pathogènes diffèrent entre les deux études. Ceci n'est qu'une illustration de l'hétérogénéité des données obtenues à travers les études que nous avons citées.

Afin de caractériser l'étendue ou, au contraire, la spécificité du rôle des miARN dans la réponse mise en place suite à l'infection, il convient d'infecter les cellules d'un même échantillon à l'aide de plusieurs pathogènes. Cette approche a été utilisée récemment dans une étude à laquelle j'ai participé (Siddle et al., 2015, voir Annexe 1). Les résultats de ce travail montrent que les modifications d'expression des miARN suite à l'infection par diverses bactéries intracellulaires peuvent être spécifiques, tant au niveau de la nature des transcrits que de l'intensité du changement de leur expression. Il a ainsi pu être possible de définir un groupe de 49 miARN correspondant au cœur de la réponse antibactérienne. Les profils d'expression d'autres miARN se sont révélés différents selon le type d'agent

infectieux utilisé pour la stimulation des cellules dendritiques, suggérant que ces transcrits font partie d'une réponse immunitaire plus spécifique.

Il serait aussi possible de définir si les réponses transcriptionnelles que nous avons mises en évidence sont restreintes aux cellules dendritiques ou si, au contraire, elles sont retrouvées dans d'autres types cellulaires de l'immunité. Pour ce faire, il serait envisageable d'appliquer le même type d'approche *ex vivo* que celle que nous avons utilisé en considérant d'autres types cellulaires. Les comparaisons les plus évidentes seraient à faire avec les macrophages, ce qui permettrait d'estimer la variabilité de réponse immunitaire parmi les cellules présentatrices d'antigènes, et les lymphocytes T, pour caractériser les similarités et les différences entre la réponse immunitaire innée et l'adaptative. Ces analyses permettraient d'estimer les types cellulaires de l'immunité présentant des différences de régulation de l'expression des miARN et d'améliorer nos connaissances sur la participation de ces transcrits à la réponse mise en place suite à l'infection.

## 7. Rôle biologique des miARN

Notre analyse identifie des différences d'expression portant sur un grand nombre de miARN, laissant supposer que ceux-ci jouent un rôle dans la réponse des cellules dendritiques à l'infection. Si certains de ces miARN, comme miR-155, ont été largement étudiés dans l'immunité (He et al., 2014; Li and Shi, 2013; Montagner et al., 2013; O'Neill et al., 2011), ils en sont d'autres tels que les miR-29 pour lesquels des mécanismes moléculaires commencent tout juste à être identifiés (Liston et al., 2012; Ma et al., 2011; Sharbati et al., 2011). Nous rapportons aussi des différences d'expression pour miR-630 et miR-339-3p, qui n'avaient jusque-là jamais été identifiés comme impliqués dans le processus immunitaire. Des études fonctionnelles permettraient donc de mieux comprendre le rôle que jouent ces régulateurs de l'expression génique dans la réponse mise en place suite à l'infection.

Nous avons par exemple réalisé des expériences de perte et gain de fonction en utilisant miR-29 comme modèle d'étude. Cependant, notre analyse est limitée. En effet, si nous trouvons un enrichissement en cibles prédites dans les gènes dérégulés suite à la surexpression de miR-29a, nous n'apportons pas la preuve d'une régulation directe de la quantité de ces transcrit par le miARN. De plus, les approches transcriptionnelles ne permettent pas d'identifier l'ensemble des mécanismes de régulation impliquant les miARN. Notre étude portant sur CXCL10, une cible prédite de miR-29, en est une illustration. En

effet, nous n'avons pas trouvé de variation significative de la quantité du transcrit correspondant lors de la perturbation de l'expression de miR-29 alors que nous avons observé des différences dans la quantité de protéine sécrétée. Là encore, il reste à démontrer que ces variations sont dues à la fixation du miARN sur les sites complémentaires présents sur l'ARNm. Pour établir ce lien, il serait par exemple possible d'effectuer des expériences utilisant un gène rapporteur. Une combinaison d'approches transcriptomiques, protéiques, ainsi que l'utilisation d'outils de biologie moléculaire permettraient donc de mettre en évidence les mécanismes moléculaires régulés par les miARN.

Comme nous venons de le mentionner, miR-29 régule la sécrétion de CXCL10 par les cellules dendritiques infectées par MTB. Cette cytokine est impliquée dans le recrutement des lymphocytes Th1 (Giacomini et al., 2006; Lande et al., 2003) qui sont particulièrement importants pour le contrôle de la croissance de la mycobactérie et la survie de l'hôte (Caruso et al., 1999; Flory et al., 1992; Scanga et al., 2000). Ainsi, l'implication de miR-29 dans des mécanismes moléculaires au sein des cellules dendritiques pourrait avoir des conséquences sur un large spectre de fonctions immunitaires. Ajouter au type d'approche que nous avons utilisé l'étude des propriétés plus complexes des cellules dendritiques comme la maturation, la phagocytose, la présentation de l'antigène, le recrutement d'autres types cellulaires ou encore la migration permettrait d'obtenir une vue d'ensemble du rôle des miARN dans la régulation de mécanismes cellulaires mis en jeu au cours de l'infection.

Il a récemment été montré que les cellules pouvaient interagir entre elles par l'intermédiaire d'exosomes ou de microvésicules contenant des ANRm et des miARN (Valadi et al., 2007; Xu et al., 2013). Ce dialogue peut être restreint au microenvironnement dans lequel les cellules évoluent ou se faire sur une longue portée si ces vésicules sont transportées par les fluides corporels comme le sang. La présence des transcrits dans ces particules extracellulaires n'est pas aléatoire mais résulte d'un tri sélectif (Guduric-Fuchs et al., 2012). De plus, les miARN transportés par les exosomes sont fonctionnels dans les cellules intégrant ces vésicules (Montecalvo et al., 2012). Les cellules dendritiques utilisent ces dérivés de la voie d'endocytose pour communiquer entre elles (Montecalvo et al., 2012), ce qui laisserait supposer que certains des miARN que nous avons identifiés comme impliqués dans la réponse de ces cellules à l'infection par MTB pourraient avoir un effet non autonome cellulaire. Parmi ces miARN, miR-29 a été rapporté comme pouvant faire partie des transcrits localisés dans des microvésicules (Fabbri et al., 2012). Il est donc tentant de

spéculer que ce miARN induit suite à l'infection des cellules dendritiques par MTB peut participer à la régulation de fonctions d'autres cellules. Par exemple, 16 des cibles prédites de miR-29 sont impliquées dans la production du collagène. On peut donc émettre l'hypothèse que, suite à la stimulation par l'agent infectieux, la quantité de miR-29 augmente dans les microvésicules produites par les cellules dendritiques. Si ces particules sont incorporées par les fibroblastes, miR-29 pourrait participer à l'inhibition de la production de matrice extracellulaire par ces cellules, facilitant ainsi la migration des cellules dendritiques, propriété qui a été démontrée comme importante dans la réponse à la tuberculose (Curtis et al., 2015). Si cette idée est très spéculative, elle illustre les perspectives qu'apporte l'identification de ce nouveau type d'interaction entre cellules dans la compréhension des mécanismes immunitaires mobilisés au cours de l'infection.

## 8. Introgression adaptative et contrôle génétique de l'expression des gènes

La combinaison de nos deux approches ouvre sur le rôle qu'aurait pu jouer l'introgression de variations génétiques provenant de génomes archaïques dans les régions intergéniques du génome des populations non africaines modernes. De forts taux d'ascendance néandertalienne sont retrouvés dans ces régions (Sankararaman et al., 2014). De plus, certains haplotypes présents dans les populations non africaines à fréquences modérées ainsi que chez les hommes archaïques portent des SNP non-codants, comme c'est le cas pour le gène de la dystrophine (Sankararaman et al., 2014; Yotova et al., 2011; Zietkiewicz et al., 2003). Les études d'associations entre variations génétiques et diversité phénotypique ont identifié de nombreux polymorphismes non-codants qui sont souvent liés à des différences d'expressions géniques (voir paragraphe suivant). De plus, les régions non-codantes seraient impliquées dans l'adaptation des organismes à leur environnement (voir la section 1.3 des conclusions et perspectives). On peut donc supposer qu'une partie de l'adaptation des populations actuelles à leur environnement pathogénique s'est faite grâce à l'introgression d'allèles archaïques dans les séquences régulatrices de l'expression des gènes impliqués dans l'immunité innée.

Mettre en correspondance des données de contrôle génétique de l'expression génique et la liste d'allèles présentant un fort taux d'ascendance néandertalienne permettrait de tester cette hypothèse. Ces dernières années, nombre d'analyses basées sur un ensemble de matériels biologiques variés (Cheung et al., 2005; Dimas et al., 2009; Fairfax et al., 2014; Han et al., 2015; Hulur et al., 2015; Kwan et al., 2009; Lee et al., 2014; Naranbhai et al.,

2015; Stranger et al., 2007; Vockley et al., 2015; Westra et al., 2015) ont mis en évidence l'étendue du contrôle génétique de l'expression génique ainsi que le fait qu'il puisse être spécifique d'un tissu ou d'un type cellulaire particulier. Cependant, peu de travaux se sont attelés à identifier le degré de variabilité d'un tel contrôle entre différentes populations humaines (Hartford et al., 2009; Spielman et al., 2007; Stranger et al., 2012). L'hypothèse d'une introgression adaptative d'allèles archaïques dans les régions régulatrices d'individus d'origines ethniques particulières impliquerait que ces variations pourraient participer à des différences d'expression génique inter-populationnelles. Cela souligne l'importance de développer des approches visant à identifier des eQTL spécifiques à différentes populations humaines dans un contexte large ou, dans le cadre de l'immunité, dans un contexte infectieux. De telles analyses permettraient de mieux comprendre l'origine des variabilités de réponse immunitaires entre populations ainsi que leurs mécanismes moléculaires.

## 9. Conclusion générale

Le travail présenté dans cette thèse contribue à améliorer notre compréhension des causes génétiques des différences interindividuelles de réponse immunitaire innée qui sont observées chez l'Homme. Nos résultats apportent de nouvelles informations quant à l'histoire évolutive des gènes de l'immunité innée, traduisant de l'importance fonctionnelle de certains, de la participation d'autres à l'adaptation des populations humaines à leur environnement, et du maintien d'allèles d'origine archaïque qui auraient pu conférer un avantage sélectif à des groupes d'individus. De façon complémentaire, nous caractérisons l'implication des miARN dans la réponse immunitaire mise en place suite à l'infection par *Mycobacterium tuberculosis*, montrons leur rôle dans le contrôle de l'expression génique au cours de cette réponse, caractérisons le contrôle génétique de leur expression et estimons les effets de l'infection sur les interactions entre génoype et phénotype. L'ensemble des variations génétiques identifiées au cours de cette thèse définit un groupe de régions génomiques dont l'étude fonctionnelle permettra de révéler certains des mécanismes moléculaires et cellulaires qui participent aux différences de susceptibilité aux maladies infectieuses.

# Bibliographie

1000 Genomes Project Consortium, 2012. An integrated map of genetic variation from 1,092 human genomes. Nature 491, 56–65. doi:10.1038/nature11632

1000 Genomes Project Consortium, 2010. A map of human genome variation from population-scale sequencing. Nature 467, 1061–1073. doi:10.1038/nature09534

Abel, L., Alcaïs, A., Schurr, E., 2014. The dissection of complex susceptibility to infectious disease: bacterial, viral and parasitic infections. Curr. Opin. Immunol. 30, 72–78. doi:10.1016/j.coi.2014.07.002

Abel, L., Casanova, J.L., 2000. Genetic predisposition to clinical tuberculosis: bridging the gap between simple and complex inheritance. Am. J. Hum. Genet. 67, 274–277. doi:10.1086/303033

Abi-Rached, L., Jobin, M.J., Kulkarni, S., McWhinnie, A., Dalva, K., Gragert, L., Babrzadeh, F., Gharizadeh, B., Luo, M., Plummer, F.A., Kimani, J., Carrington, M., Middleton, D., Rajalingam, R., Beksac, M., Marsh, S.G.E., Maiers, M., Guethlein, L.A., Tavoularis, S., Little, A.-M., Green, R.E., Norman, P.J., Parham, P., 2011. The shaping of modern human immune systems by multiregional admixture with archaic humans. Science 334, 89–94. doi:10.1126/science.1209202

Ablasser, A., Bauernfeind, F., Hartmann, G., Latz, E., Fitzgerald, K.A., Hornung, V., 2009. RIG-I-dependent sensing of poly(dA:dT) through the induction of an RNA polymerase III-transcribed RNA intermediate. Nat. Immunol. 10, 1065–1072. doi:10.1038/ni.1779

Akira, S., Takeda, K., 2004. Toll-like receptor signalling. Nat. Rev. Immunol. 4, 499–511. doi:10.1038/nri1391

Akira, S., Uematsu, S., Takeuchi, O., 2006. Pathogen recognition and innate immunity. Cell 124, 783–801. doi:10.1016/j.cell.2006.02.015

Alcais, A., Abel, L., Casanova, J.L., 2009. Human genetics of infectious diseases: between proof of principle and paradigm. The Journal of clinical investigation 119, 2506–14. doi:10.1172/JCI38111

Alcaïs, A., Fieschi, C., Abel, L., Casanova, J.L., 2005. Tuberculosis in children and adults: two distinct genetic diseases. J. Exp. Med. 202, 1617–1621. doi:10.1084/jem.20052302

Alcaïs, A., Quintana-Murci, L., Thaler, D.S., Schurr, E., Abel, L., Casanova, J.L., 2010. Life-threatening infectious diseases of childhood: single-gene inborn errors of immunity? Ann. N. Y. Acad. Sci. 1214, 18–33. doi:10.1111/j.1749-6632.2010.05834.x

Aldred, P.M.R., Hollox, E.J., Armour, J.A.L., 2005. Copy number polymorphism and expression level variation of the human alpha-defensin genes DEFA1 and DEFA3. Hum. Mol. Genet. 14, 2045–2052. doi:10.1093/hmg/ddi209

Al-Herz, W., Bousfiha, A., Casanova, J.L., Chatila, T., Conley, M.E., Cunningham-Rundles, C., Etzioni, A., Franco, J.L., Gaspar, H.B., Holland, S.M., Klein, C., Nonoyama, S., Ochs, H.D., Oksenhendler, E., Picard, C., Puck, J.M., Sullivan, K., Tang, M.L., 2014. Primary immunodeficiency diseases: an update on the classification from the international union of immunological societies expert committee for primary immunodeficiency. Frontiers in immunology 5, 162. doi:10.3389/fimmu.2014.00162

Allison, A.C., 1954. Protection afforded by sickle-cell trait against subtertian malareal infection. British medical journal 1, 290–4.

Altare, F., Durandy, A., Lammas, D., Emile, J.F., Lamhamedi, S., Le Deist, F., Drysdale, P., Jouanguy, E., Döffinger, R., Bernaudin, F., Jeppsson, O., Gollob, J.A., Meinl, E., Segal, A.W., Fischer, A., Kumararatne, D., Casanova, J.L., 1998a. Impairment of mycobacterial immunity in human interleukin-12 receptor deficiency. Science 280, 1432–1435.

Altare, F., Lammas, D., Revy, P., Jouanguy, E., Döffinger, R., Lamhamedi, S., Drysdale, P., Scheel-Toellner, D., Girdlestone, J., Darbyshire, P., Wadhwa, M., Dockrell, H., Salmon, M., Fischer, A., Durandy, A., Casanova, J.L., Kumararatne, D.S., 1998b. Inherited interleukin 12 deficiency in a child with bacille Calmette-Guérin and Salmonella enteritidis disseminated infection. J. Clin. Invest. 102, 2035–2040. doi:10.1172/JCI4950

Ambros, V., 2004. The functions of animal microRNAs. Nature 431, 350–355. doi:10.1038/nature02871

Ambros, V., Bartel, B., Bartel, D.P., Burge, C.B., Carrington, J.C., Chen, X., Dreyfuss, G., Eddy, S.R., Griffiths-Jones, S., Marshall, M., Matzke, M., Ruvkun, G., Tuschl, T., 2003. A uniform system for microRNA annotation. RNA 9, 277–279.

Anderson, K.V., Jürgens, G., Nüsslein-Volhard, C., 1985. Establishment of dorsal-ventral polarity in the Drosophila embryo: genetic studies on the role of the Toll gene product. Cell 42, 779–789.

Angulo, I., Vadas, O., Garçon, F., Banham-Hall, E., Plagnol, V., Leahy, T.R., Baxendale, H., Coulter, T., Curtis, J., Wu, C., Blake-Palmer, K., Perisic, O., Smyth, D., Maes, M., Fiddler, C., Juss, J., Cilliers, D., Markelj, G., Chandra, A., Farmer, G., Kielkowska, A., Clark, J., Kracker, S., Debré, M., Picard, C., Pellier, I., Jabado, N., Morris, J.A., Barcenas-Morales, G., Fischer, A., Stephens, L., Hawkins, P., Barrett, J.C., Abinun, M., Clatworthy, M., Durandy, A., Doffinger, R., Chilvers, E.R., Cant, A.J., Kumararatne, D., Okkenhaug, K., Williams, R.L., Condliffe, A., Nejentsev, S., 2013. Phosphoinositide 3-kinase δ gene mutation predisposes to respiratory infection and airway damage. Science 342, 866–871. doi:10.1126/science.1243292

Ariumi, Y., 2014. Multiple functions of DDX3 RNA helicase in gene regulation, tumorigenesis, and viral infection. Front Genet 5, 423. doi:10.3389/fgene.2014.00423

Badano, J.L., Katsanis, N., 2002. Beyond Mendel: an evolving view of human genetic disease transmission. Nat. Rev. Genet. 3, 779–789. doi:10.1038/nrg910

Baek, D., Villén, J., Shin, C., Camargo, F.D., Gygi, S.P., Bartel, D.P., 2008. The impact of microRNAs on protein output. Nature 455, 64–71. doi:10.1038/nature07242

Baghdadi, J. El, Orlova, M., Alter, A., Ranque, B., Chentoufi, M., Lazrak, F., Archane, M.I., Casanova, J.L., Benslimane, A., Schurr, E., Abel, L., 2006. An autosomal dominant major gene confers predisposition to pulmonary tuberculosis in adults. J. Exp. Med. 203, 1679–1684. doi:10.1084/jem.20060269

Barreiro, L.B., Ben-Ali, M., Quach, H., Laval, G., Patin, E., Pickrell, J.K., Bouchier, C., Tichit, M., Neyrolles, O., Gicquel, B., Kidd, J.R., Kidd, K.K., Alcaïs, A., Ragimbeau, J., Pellegrini, S., Abel, L., Casanova, J.L., Quintana-Murci, L., 2009. Evolutionary dynamics of human Toll-like receptors and their different contributions to host defense. PLoS Genet. 5, e1000562. doi:10.1371/journal.pgen.1000562

Barreiro, L.B., Laval, G., Quach, H., Patin, E., Quintana-Murci, L., 2008. Natural selection has driven population differentiation in modern humans. Nature genetics 40, 340–5. doi:10.1038/ng.78

Barreiro, L.B., Neyrolles, O., Babb, C.L., Tailleux, L., Quach, H., McElreavey, K., van Helden, P.D., Hoal, E.G., Gicquel, B., Quintana-Murci, L., 2006. Promoter variation in the DC-SIGN-encoding gene CD209 is associated with tuberculosis. PLoS Med. 3, e20. doi:10.1371/journal.pmed.0030020

Barreiro, L.B., Quintana-Murci, L., 2010. From evolutionary genetics to human immunology: how selection shapes host defence genes. Nat. Rev. Genet. 11, 17–30. doi:10.1038/nrg2698

Barreiro, L.B., Tailleux, L., Pai, A.A., Gicquel, B., Marioni, J.C., Gilad, Y., 2012. Deciphering the genetic architecture of variation in the immune response to Mycobacterium tuberculosis infection. Proc. Natl. Acad. Sci. U.S.A. 109, 1204–1209. doi:10.1073/pnas.1115761109

Bartel, D.P., Chen, C.-Z., 2004. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. Nat. Rev. Genet. 5, 396–400. doi:10.1038/nrg1328

Baskerville, S., Bartel, D.P., 2005. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. RNA 11, 241–247. doi:10.1261/rna.7240905

Baum, A., García-Sastre, A., 2010. Induction of type I interferon by RNA viruses: cellular receptors and their substrates. Amino Acids 38, 1283–1299. doi:10.1007/s00726-009-0374-0

Bellamy, R., Ruwende, C., Corrah, T., McAdam, K.P., Whittle, H.C., Hill, A.V., 1998. Variations in the NRAMP1 gene and susceptibility to tuberculosis in West Africans. N. Engl. J. Med. 338, 640–644. doi:10.1056/NEJM199803053381002

Belton, J.-M., McCord, R.P., Gibcus, J.H., Naumova, N., Zhan, Y., Dekker, J., 2012. Hi-C: a comprehensive technique to capture the conformation of genomes. Methods 58, 268–276. doi:10.1016/j.ymeth.2012.05.001

Ben-Ali, M., Corre, B., Manry, J., Barreiro, L.B., Quach, H., Boniotto, M., Pellegrini, S., Quintana-Murci, L., 2011. Functional characterization of naturally occurring genetic variants in the human TLR1-2-6 gene family. Hum. Mutat. 32, 643–652. doi:10.1002/humu.21486

Berezikov, E., Chung, W.-J., Willis, J., Cuppen, E., Lai, E.C., 2007. Mammalian mirtron genes. Mol. Cell 28, 328–336. doi:10.1016/j.molcel.2007.09.028

Bernig, T., Taylor, J.G., Foster, C.B., Staats, B., Yeager, M., Chanock, S.J., 2004. Sequence analysis of the mannose-binding lectin (MBL2) gene reveals a high degree of heterozygosity with evidence of selection. Genes Immun. 5, 461–476. doi:10.1038/sj.gene.6364116

Bersaglieri, T., Sabeti, P.C., Patterson, N., Vanderploeg, T., Schaffner, S.F., Drake, J.A., Rhodes, M., Reich, D.E., Hirschhorn, J.N., 2004. Genetic signatures of strong recent positive selection at the lactase gene. Am. J. Hum. Genet. 74, 1111–1120. doi:10.1086/421051

Beutler, B., Jiang, Z., Georgel, P., Crozat, K., Croker, B., Rutschmann, S., Du, X., Hoebe, K., 2006. GENETIC ANALYSIS OF HOST RESISTANCE: Toll-Like Receptor Signaling and Immunity at Large. Annual Review of Immunology 24, 353–389. doi:10.1146/annurev.immunol.24.021605.090552

Bhatia, G., Patterson, N., Pasaniuc, B., Zaitlen, N., Genovese, G., Pollack, S., Mallick, S., Myers, S., Tandon, A., Spencer, C., Palmer, C.D., Adeyemo, A.A., Akylbekova, E.L., Cupples, L.A., Divers, J., Fornage, M., Kao, W.H.L., Lange, L., Li, M., Musani, S., Mychaleckyj, J.C., Ogunniyi, A., Papanicolaou, G., Rotimi, C.N., Rotter, J.I., Ruczinski, I., Salako, B., Siscovick, D.S., Tayo, B.O., Yang, Q., McCarroll, S., Sabeti, P., Lettre, G., De Jager, P., Hirschhorn, J., Zhu, X., Cooper, R., Reich, D., Wilson, J.G., Price, A.L., 2011. Genome-wide comparison of African-ancestry populations from CARe and other cohorts reveals signals of natural selection. Am. J. Hum. Genet. 89, 368–381. doi:10.1016/j.ajhg.2011.07.025

Blasius, A.L., Beutler, B., 2010. Intracellular toll-like receptors. Immunity 32, 305–315. doi:10.1016/j.immuni.2010.03.012

Bogunovic, D., Byun, M., Durfee, L.A., Abhyankar, A., Sanal, O., Mansouri, D., Salem, S., Radovanovic, I., Grant, A.V., Adimi, P., Mansouri, N., Okada, S., Bryant, V.L., Kong, X.-F., Kreins, A., Velez, M.M., Boisson, B., Khalilzadeh, S., Ozcelik, U., Darazam, I.A., Schoggins, J.W., Rice, C.M., Al-Muhsen, S., Behr, M., Vogt, G., Puel, A., Bustamante, J., Gros, P., Huibregtse, J.M., Abel, L., Boisson-Dupuis, S., Casanova, J.-L., 2012. Mycobacterial disease and impaired IFN-γ immunity in humans with inherited ISG15 deficiency. Science 337, 1684–1688. doi:10.1126/science.1224026

Boisson-Dupuis, S., Kong, X.-F., Okada, S., Cypowyj, S., Puel, A., Abel, L., Casanova, J.-L., 2012. Inborn errors of human STAT1: allelic heterogeneity governs the diversity of immunological and infectious phenotypes. Curr. Opin. Immunol. 24, 364–378. doi:10.1016/j.coi.2012.04.011

Botstein, D., Risch, N., 2003. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat. Genet. 33 Suppl, 228–237. doi:10.1038/ng1090

Bousfiha, A.A., Jeddane, L., Ailal, F., Benhsaien, I., Mahlaoui, N., Casanova, J.L., Abel, L., 2013. Primary immunodeficiency diseases worldwide: more common than generally thought. Journal of clinical immunology 33, 1–7. doi:10.1007/s10875-012-9751-7

Braff, M.H., Bardan, A., Nizet, V., Gallo, R.L., 2005. Cutaneous defense mechanisms by antimicrobial peptides. J. Invest. Dermatol. 125, 9–13. doi:10.1111/j.0022-202X.2004.23587.x

Brunette, R.L., Young, J.M., Whitley, D.G., Brodsky, I.E., Malik, H.S., Stetson, D.B., 2012. Extensive evolutionary and functional diversity among mammalian AIM2-like receptors. J. Exp. Med. 209, 1969–1983. doi:10.1084/jem.20121960

Burdette, D.L., Monroe, K.M., Sotelo-Troha, K., Iwig, J.S., Eckert, B., Hyodo, M., Hayakawa, Y., Vance, R.E., 2011. STING is a direct innate immune sensor of cyclic di-GMP. Nature 478, 515–518. doi:10.1038/nature10429

Burdette, D.L., Vance, R.E., 2013. STING and the innate immune response to nucleic acids in the cytosol. Nat. Immunol. 14, 19–26. doi:10.1038/ni.2491

Burns, S.O., Plagnol, V., Gutierrez, B.M., Zahrani, D. Al, Curtis, J., Gaspar, M., Hassan, A., Jones, A.M., Malone, M., Rampling, D., McLatchie, A., Doffinger, R., Gilmour, K.C., Henriquez, F., Thrasher, A.J., Gaspar, H.B., Nejentsev, S., 2014. Immunodeficiency and disseminated mycobacterial infection associated with homozygous nonsense mutation of IKKβ. J. Allergy Clin. Immunol. 134, 215–218. doi:10.1016/j.jaci.2013.12.1093

Bustamante, C.D., Fledel-Alon, A., Williamson, S., Nielsen, R., Hubisz, M.T., Glanowski, S., Tanenbaum, D.M., White, T.J., Sninsky, J.J., Hernandez, R.D., Civello, D., Adams, M.D., Cargill, M., Clark, A.G., 2005. Natural selection on protein-coding genes in the human genome. Nature 437, 1153–7. doi:10.1038/nature04240

Bustamante, J., Arias, A.A., Vogt, G., Picard, C., Galicia, L.B., Prando, C., Grant, A.V., Marchal, C.C., Hubeau, M., Chapgier, A., de Beaucoudrey, L., Puel, A., Feinberg, J., Valinetz, E., Jannière, L., Besse, C., Boland, A., Brisseau, J.-M., Blanche, S., Lortholary, O., Fieschi, C., Emile, J.-F., Boisson-Dupuis, S., Al-Muhsen, S., Woda, B., Newburger, P.E., Condino-Neto, A., Dinauer, M.C., Abel, L., Casanova, J.-L., 2011. Germline CYBB mutations that selectively affect macrophages in kindreds with X-linked predisposition to tuberculous mycobacterial disease. Nat. Immunol. 12, 213–221. doi:10.1038/ni.1992

Bustamante, J., Boisson-Dupuis, S., Abel, L., Casanova, J.-L., 2014. Mendelian susceptibility to mycobacterial disease: genetic, immunological, and clinical features of inborn errors of IFN-γ immunity. Semin. Immunol. 26, 454–470. doi:10.1016/j.smim.2014.09.008

Butler, J. a. V., Smith, K.A., 1950. Degradation of deoxyribonucleic acid by free radicals. Nature 165, 847–848.

Byun, M., Abhyankar, A., Lelarge, V., Plancoulaine, S., Palanduz, A., Telhan, L., Boisson, B., Picard, C., Dewell, S., Zhao, C., Jouanguy, E., Feske, S., Abel, L., Casanova, J.L., 2010. Whole-exome sequencing-based discovery of STIM1 deficiency in a child with fatal classic Kaposi sarcoma. The Journal of experimental medicine 207, 2307–12. doi:10.1084/jem.20101597

Cagliani, R., Forni, D., Biasin, M., Comabella, M., Guerini, F.R., Riva, S., Pozzoli, U., Agliardi, C., Caputo, D., Malhotra, S., Montalban, X., Bresolin, N., Clerici, M., Sironi, M., 2014. Ancient and recent selective pressures shaped genetic diversity at AIM2-like nucleic acid sensors. Genome Biol Evol 6, 830–845. doi:10.1093/gbe/evu066

Cagliani, R., Fumagalli, M., Riva, S., Pozzoli, U., Comi, G.P., Menozzi, G., Bresolin, N., Sironi, M., 2008. The signature of long-standing balancing selection at the human defensin beta-1 promoter. Genome Biol. 9, R143. doi:10.1186/gb-2008-9-9-r143

Cairns, J., 1998. Matters of Life and Death: Perspectives on Public Health, Molecular Biology, Cancer, and the Prospects for the Human Race. Princeton University Press.

Caligiuri, M.A., 2008. Human natural killer cells. Blood 112, 461–9. doi:10.1182/blood-2007-09-077438

Calmette, A., Guérin, C., Weill-Hallé, B., Nègre, L., Boquet, A., 1925. Essai de prémunition par le BCG contre l'infection tuberculeuse de l'homme et des animaux. Bull. Acad. méd. 93, 78.

Campbell, C.D., Chong, J.X., Malig, M., Ko, A., Dumont, B.L., Han, L., Vives, L., O'Roak, B.J., Sudmant, P.H., Shendure, J., Abney, M., Ober, C., Eichler, E.E., 2012. Estimating the human mutation rate using autozygosity in a founder population. Nature Genetics 44, 1277–1281. doi:10.1038/ng.2418

Campbell, M.C., Tishkoff, S.A., 2008. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. Annu Rev Genomics Hum Genet 9, 403–433. doi:10.1146/annurev.genom.9.081307.164258

Cann, R.L., Stoneking, M., Wilson, A.C., 1987. Mitochondrial DNA and human evolution. Nature 325, 31–36. doi:10.1038/325031a0

Cao, W., Zhang, L., Rosen, D.B., Bover, L., Watanabe, G., Bao, M., Lanier, L.L., Liu, Y.-J., 2007. BDCA2/Fc epsilon RI gamma complex signals through a novel BCR-like pathway in human plasmacytoid dendritic cells. PLoS Biol. 5, e248. doi:10.1371/journal.pbio.0050248

Caplan, J., Padmanabhan, M., Dinesh-Kumar, S.P., 2008. Plant NB-LRR immune receptors: from recognition to transcriptional reprogramming. Cell Host Microbe 3, 126–135. doi:10.1016/j.chom.2008.02.010

Cappellini, M.D., Fiorelli, G., 2008. Glucose-6-phosphate dehydrogenase deficiency. Lancet 371, 64–74. doi:10.1016/S0140-6736(08)60073-2

Carlson, C.S., Thomas, D.J., Eberle, M.A., Swanson, J.E., Livingston, R.J., Rieder, M.J., Nickerson, D.A., 2005. Genomic regions exhibiting positive selection identified from dense genotype data. Genome research 15, 1553–65. doi:10.1101/gr.4326505

Carter, R., Mendis, K.N., 2002. Evolutionary and historical aspects of the burden of malaria. Clinical microbiology reviews 15, 564–94.

Caruso, A.M., Serbina, N., Klein, E., Triebold, K., Bloom, B.R., Flynn, J.L., 1999. Mice deficient in CD4 T cells have only transiently diminished levels of IFN-gamma, yet succumb to tuberculosis. J. Immunol. 162, 5407–5416.

Casanova, J.L., Abel, L., 2004. The human model: a genetic dissection of immunity to infection in natural conditions. Nature reviews. Immunology 4, 55–66. doi:10.1038/nri1264

Casanova, J.-L., Abel, L., 2004. Human Mannose-binding Lectin in Immunity: Friend, Foe, or Both? J. Exp. Med. 199, 1295–1299. doi:10.1084/jem.20040537

Casanova, J.L., Abel, L., 2002. Genetic dissection of immunity to mycobacteria: the human model. Annu. Rev. Immunol. 20, 581–620. doi:10.1146/annurev.immunol.20.081501.125851

Casanova, J.-L., Abel, L., Quintana-Murci, L., 2013. Immunology taught by human genetics. Cold Spring Harb. Symp. Quant. Biol. 78, 157–172. doi:10.1101/sqb.2013.78.019968

Casrouge, A., Zhang, S.Y., Eidenschenk, C., Jouanguy, E., Puel, A., Yang, K., Alcais, A., Picard, C., Mahfoufi, N., Nicolas, N., Lorenzo, L., Plancoulaine, S., Senechal, B., Geissmann, F., Tabeta, K., Hoebe, K., Du, X., Miller, R.L., Heron, B., Mignot, C., de Villemeur, T.B., Lebon, P., Dulac, O., Rozenberg, F., Beutler, B., Tardieu, M., Abel, L., Casanova, J.L., 2006. Herpes simplex virus encephalitis in human UNC-93B deficiency. Science 314, 308–12. doi:10.1126/science.1128346

Castellano, S., Parra, G., Sánchez-Quinto, F.A., Racimo, F., Kuhlwilm, M., Kircher, M., Sawyer, S., Fu, Q., Heinze, A., Nickel, B., Dabney, J., Siebauer, M., White, L., Burbano, H.A., Renaud, G., Stenzel, U., Lalueza-Fox, C., de la Rasilla, M., Rosas, A., Rudan, P., Brajković, D., Kucan, Ž., Gušic, I., Shunkov, M.V., Derevianko, A.P., Viola, B., Meyer, M., Kelso, J., Andrés, A.M., Pääbo, S., 2014. Patterns of coding variation in the complete exomes of three Neandertals. Proc. Natl. Acad. Sci. U.S.A. 111, 6666–6671. doi:10.1073/pnas.1405138111

Cavalli-Sforza, L.L., 1966. Population structure and human evolution. Proc. R. Soc. Lond., B, Biol. Sci. 164, 362–379.

Cavalli-Sforza, L.L., Feldman, M.W., 2003. The application of molecular genetic approaches to the study of human evolution. Nat. Genet. 33 Suppl, 266–275. doi:10.1038/ng1113

Chang, H.-T., Li, S.-C., Ho, M.-R., Pan, H.-W., Ger, L.-P., Hu, L.-Y., Yu, S.-Y., Li, W.-H., Tsai, K.-W., 2012. Comprehensive analysis of microRNAs in breast cancer. BMC Genomics 13 Suppl 7, S18. doi:10.1186/1471-2164-13-S7-S18

Chapman, S.J., Hill, A.V.S., 2012. Human genetic susceptibility to infectious disease. Nat Rev Genet 13, 175–188. doi:10.1038/nrg3114

Charlesworth, D., 2006. Balancing selection and its effects on sequences in nearby genome regions. PLoS Genet. 2, e64. doi:10.1371/journal.pgen.0020064

Chen, H., Patterson, N., Reich, D., 2010. Population differentiation as a test for selective sweeps. Genome Res. 20, 393–402. doi:10.1101/gr.100545.109

Chen, Y.S., Torroni, A., Excoffier, L., Santachiara-Benerecetti, A.S., Wallace, D.C., 1995. Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. American journal of human genetics 57, 133–49.

Cheung, V.G., Spielman, R.S., Ewens, K.G., Weber, T.M., Morley, M., Burdick, J.T., 2005. Mapping determinants of human gene expression by regional and genome-wide association. Nature 437, 1365–1369. doi:10.1038/nature04244

Chiang, H.R., Schoenfeld, L.W., Ruby, J.G., Auyeung, V.C., Spies, N., Baek, D., Johnston, W.K., Russ, C., Luo, S., Babiarz, J.E., Blelloch, R., Schroth, G.P., Nusbaum, C., Bartel, D.P., 2010. Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. Genes Dev. 24, 992–1009. doi:10.1101/gad.1884710

Chimusa, E.R., Zaitlen, N., Daya, M., Möller, M., van Helden, P.D., Mulder, N.J., Price, A.L., Hoal, E.G., 2014. Genome-wide association study of ancestry-specific TB risk in the South African Coloured population. Hum Mol Genet 23, 796–809. doi:10.1093/hmg/ddt462

Chiu, Y.-H., Macmillan, J.B., Chen, Z.J., 2009. RNA polymerase III detects cytosolic DNA and induces type I interferons through the RIG-I pathway. Cell 138, 576–591. doi:10.1016/j.cell.2009.06.015

Choremis, C., Kattamis, C.A., Kyriazakou, M., Gavriilidou, E., 1966. Viral hepatitis in G.-6-P.D. deficiency. Lancet 1, 269–70.

Chuang, T.H., Ulevitch, R.J., 2000. Cloning and characterization of a sub-family of human toll-like receptors: hTLR7, hTLR8 and hTLR9. Eur. Cytokine Netw. 11, 372–378.

Chuang, T., Ulevitch, R.J., 2001. Identification of hTLR10: a novel human Toll-like receptor preferentially expressed in immune cells. Biochim. Biophys. Acta 1518, 157–161.

Chun, E.H., Gonzales, L., Lewis, F.S., Jones, J., Rutman, R.J., 1969. Differences in the in vivo alkylation and cross-linking of nitrogen mustard-sensitive and -resistant lines of Lettré-Ehrlich asites tumors. Cancer Res. 29, 1184–1194.

Clark, A.G., Glanowski, S., Nielsen, R., Thomas, P.D., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civello, D., Lu, F., Murphy, B., Ferriera, S., Wang, G., Zheng, X., White, T.J., Sninsky, J.J., Adams, M.D., Cargill, M., 2003. Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. Science 302, 1960–1963. doi:10.1126/science.1088821

Cloonan, N., Wani, S., Xu, Q., Gu, J., Lea, K., Heater, S., Barbacioru, C., Steptoe, A.L., Martin, H.C., Nourbakhsh, E., Krishnan, K., Gardiner, B., Wang, X., Nones, K., Steen, J.A., Matigian, N.A., Wood, D.L., Kassahn, K.S., Waddell, N., Shepherd, J., Lee, C., Ichikawa, J., McKernan, K., Bramlett, K., Kuersten, S., Grimmond, S.M., 2011. MicroRNAs and their isomiRs function cooperatively to target common biological pathways. Genome Biol. 12, R126. doi:10.1186/gb-2011-12-12-r126

Comstock, G.W., 1978. Tuberculosis in twins: a re-analysis of the Prophit survey. Am. Rev. Respir. Dis. 117, 621–624.

Conley, M.E., Casanova, J.-L., 2014. Discovery of single-gene inborn errors of immunity by next generation sequencing. Curr. Opin. Immunol. 30, 17–23. doi:10.1016/j.coi.2014.05.004

Cross, C.E., Halliwell, B., Borish, E.T., Pryor, W.A., Ames, B.N., Saul, R.L., McCord, J.M., Harman, D., 1987. Oxygen radicals and human disease. Ann. Intern. Med. 107, 526–545.

Curtis, J., Luo, Y., Zenner, H.L., Cuchet-Lourenço, D., Wu, C., Lo, K., Maes, M., Alisaac, A., Stebbings, E., Liu, J.Z., Kopanitsa, L., Ignatyeva, O., Balabanova, Y., Nikolayevskyy, V., Baessmann, I., Thye, T., Meyer, C.G., Nürnberg, P., Horstmann, R.D., Drobniewski, F., Plagnol, V., Barrett, J.C., Nejentsev, S., 2015. Susceptibility to tuberculosis is associated with variants in the ASAP1 gene encoding a regulator of dendritic cell migration. Nat Genet 47, 523–527. doi:10.1038/ng.3248

Darwin, C., 1859. On the Origin of Species by Means of Natural Selection, Or, The Preservation of Favoured Races in the Struggle for Life. J. Murray.

Das, S., Nikolaidis, N., Goto, H., McCallister, C., Li, J., Hirano, M., Cooper, M.D., 2010. Comparative genomics and evolution of the alpha-defensin multigene family in primates. Mol. Biol. Evol. 27, 2333–2343. doi:10.1093/molbev/msq118

Daub, J.T., Hofer, T., Cutivet, E., Dupanloup, I., Quintana-Murci, L., Robinson-Rechavi, M., Excoffier, L., 2013. Evidence for polygenic adaptation to pathogens in the human genome. Mol. Biol. Evol. 30, 1544–1558. doi:10.1093/molbev/mst080

Davila, S., Hibberd, M.L., Hari Dass, R., Wong, H.E.E., Sahiratmadja, E., Bonnard, C., Alisjahbana, B., Szeszko, J.S., Balabanova, Y., Drobniewski, F., van Crevel, R., van de Vosse, E., Nejentsev, S., Ottenhoff, T.H.M., Seielstad, M., 2008. Genetic association and expression studies indicate a role of toll-like receptor 8 in pulmonary tuberculosis. PLoS Genet. 4, e1000218. doi:10.1371/journal.pgen.1000218

Deane, J.A., Bolland, S., 2006. Nucleic acid-sensing TLRs as modifiers of autoimmunity. J. Immunol. 177, 6573–6578.

Dean, M., Carrington, M., Winkler, C., Huttley, G.A., Smith, M.W., Allikmets, R., Goedert, J.J., Buchbinder, S.P., Vittinghoff, E., Gomperts, E., Donfield, S., Vlahov, D., Kaslow, R., Saah, A., Rinaldo, C., Detels, R., O'Brien, S.J., 1996. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. Science 273, 1856–62.

De Bont, R., van Larebeke, N., 2004. Endogenous DNA damage in humans: a review of quantitative data. Mutagenesis 19, 169–185.

de Jong, R., Altare, F., Haagen, I.A., Elferink, D.G., Boer, T., van Breda Vriesman, P.J., Kabel, P.J., Draaisma, J.M., van Dissel, J.T., Kroon, F.P., Casanova, J.L., Ottenhoff, T.H., 1998. Severe mycobacterial and Salmonella infections in interleukin-12 receptor-deficient patients. Science 280, 1435–1438.

Dempsey, A., Bowie, A.G., 2015. Innate immune recognition of DNA: A recent history. Virology, 60th Anniversary Issue 479–480, 146–152. doi:10.1016/j.virol.2015.03.013

Dermitzakis, E.T., 2012. Cellular genomics for complex traits. Nat. Rev. Genet. 13, 215–220. doi:10.1038/nrg3115

Dermitzakis, E.T., 2008. Regulatory variation and evolution: implications for disease. Adv. Genet. 61, 295–306. doi:10.1016/S0065-2660(07)00011-9

Dimas, A.S., Deutsch, S., Stranger, B.E., Montgomery, S.B., Borel, C., Attar-Cohen, H., Ingle, C., Beazley, C., Gutierrez Arcelus, M., Sekowska, M., Gagnebin, M., Nisbett, J., Deloukas, P., Dermitzakis, E.T., Antonarakis, S.E., 2009. Common regulatory variation impacts gene expression in a cell type-dependent manner. Science 325, 1246–1250. doi:10.1126/science.1174148

Ding, Q., Hu, Y., Jin, L., 2014a. Non-Neanderthal Origin of the HLA-DPB1*0401. J. Biol. Chem. 289, 10252–10252. doi:10.1074/jbc.L114.547505

Ding, Q., Hu, Y., Xu, S., Wang, C.-C., Li, H., Zhang, R., Yan, S., Wang, J., Jin, L., 2014b. Neanderthal origin of the haplotypes carrying the functional variant Val92Met in the MC1R in modern humans. Mol. Biol. Evol. 31, 1994–2003. doi:10.1093/molbev/msu180

Dixit, E., Boulant, S., Zhang, Y., Lee, A.S.Y., Odendall, C., Shum, B., Hacohen, N., Chen, Z.J., Whelan, S.P., Fransen, M., Nibert, M.L., Superti-Furga, G., Kagan, J.C., 2010. Peroxisomes are signaling platforms for antiviral innate immunity. Cell 141, 668–681. doi:10.1016/j.cell.2010.04.018

Dombroski, B.A., Nayak, R.R., Ewens, K.G., Ankener, W., Cheung, V.G., Spielman, R.S., 2010. Gene expression and genetic variation in response to endoplasmic reticulum stress in human cells. Am. J. Hum. Genet. 86, 719–729. doi:10.1016/j.ajhg.2010.03.017

Donnelly, R.P., Kotenko, S.V., 2010. Interferon-lambda: a new addition to an old family. J. Interferon Cytokine Res. 30, 555–564. doi:10.1089/jir.2010.0078

Drummond, R.A., Brown, G.D., 2013. Signalling C-type lectins in antimicrobial immunity. PLoS Pathog. 9, e1003417. doi:10.1371/journal.ppat.1003417

Dupuis, S., Dargemont, C., Fieschi, C., Thomassin, N., Rosenzweig, S., Harris, J., Holland, S.M., Schreiber, R.D., Casanova, J.L., 2001. Impairment of mycobacterial but not viral immunity by a germline human STAT1 mutation. Science 293, 300–303. doi:10.1126/science.1061154

Du, X., Poltorak, A., Wei, Y., Beutler, B., 2000. Three novel mammalian toll-like receptors: gene structure, expression, and evolution. Eur. Cytokine Netw. 11, 362–371.

Eilertson, K.E., Booth, J.G., Bustamante, C.D., 2012. SnIPRE: selection inference using a Poisson random effects model. PLoS Comput. Biol. 8, e1002806. doi:10.1371/journal.pcbi.1002806

Enard, D., Depaulis, F., Roest Crollius, H., 2010. Human and non-human primate genomes share hotspots of positive selection. PLoS Genet. 6, e1000840. doi:10.1371/journal.pgen.1000840

Enard, D., Messer, P.W., Petrov, D.A., 2014. Genome-wide signals of positive selection in human evolution. Genome Res. 24, 885–895. doi:10.1101/gr.164822.113

ENCODE Project Consortium, 2012. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74. doi:10.1038/nature11247

Eulalio, A., Huntzinger, E., Izaurralde, E., 2008. GW182 interaction with Argonaute is essential for miRNA-mediated translational repression and mRNA decay. Nat. Struct. Mol. Biol. 15, 346–353. doi:10.1038/nsmb.1405

Eulalio, A., Huntzinger, E., Nishihara, T., Rehwinkel, J., Fauser, M., Izaurralde, E., 2009a. Deadenylation is a widespread effect of miRNA regulation. RNA 15, 21–32. doi:10.1261/rna.1399509

Eulalio, A., Tritschler, F., Izaurralde, E., 2009b. The GW182 protein family in animal cells: new insights into domains required for miRNA-mediated gene silencing. RNA 15, 1433–1442. doi:10.1261/rna.1703809

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V.C., Foll, M., 2013. Robust demographic inference from genomic and SNP data. PLoS Genet. 9, e1003905. doi:10.1371/journal.pgen.1003905

Fabbri, M., Paone, A., Calore, F., Galli, R., Gaudio, E., Santhanam, R., Lovat, F., Fadda, P., Mao, C., Nuovo, G.J., Zanesi, N., Crawford, M., Ozer, G.H., Wernicke, D., Alder, H., Caligiuri, M.A., Nana-Sinkam, P., Perrotti, D., Croce, C.M., 2012. MicroRNAs bind to Toll-like receptors to induce prometastatic inflammatory response. Proc. Natl. Acad. Sci. U.S.A. 109, E2110–2116. doi:10.1073/pnas.1209414109

Fagny, M., Patin, E., Enard, D., Barreiro, L.B., Quintana-Murci, L., Laval, G., 2014. Exploring the Occurrence of Classic Selective Sweeps in Humans Using Whole-Genome Sequencing Data Sets. Mol Biol Evol 31, 1850–1868. doi:10.1093/molbev/msu118

Fagundes, N.J.R., Ray, N., Beaumont, M., Neuenschwander, S., Salzano, F.M., Bonatto, S.L., Excoffier, L., 2007. Statistical evaluation of alternative models of human evolution. Proc. Natl. Acad. Sci. U.S.A. 104, 17614–17619. doi:10.1073/pnas.0708280104

Fairfax, B.P., Humburg, P., Makino, S., Naranbhai, V., Wong, D., Lau, E., Jostins, L., Plant, K., Andrews, R., McGee, C., Knight, J.C., 2014. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. Science 343, 1246949. doi:10.1126/science.1246949

Fairfax, B.P., Makino, S., Radhakrishnan, J., Plant, K., Leslie, S., Dilthey, A., Ellis, P., Langford, C., Vannberg, F.O., Knight, J.C., 2012. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. Nat. Genet. 44, 502–510. doi:10.1038/ng.2205

Fay, J.C., Wu, C.I., 2000. Hitchhiking under positive Darwinian selection. Genetics 155, 1405–1413.

Fehrmann, R.S.N., Jansen, R.C., Veldink, J.H., Westra, H.-J., Arends, D., Bonder, M.J., Fu, J., Deelen, P., Groen, H.J.M., Smolonska, A., Weersma, R.K., Hofstra, R.M.W., Buurman, W.A., Rensen, S., Wolfs, M.G.M., Platteel, M., Zhernakova, A., Elbers, C.C., Festen, E.M., Trynka, G., Hofker, M.H., Saris, C.G.J., Ophoff, R.A., van den Berg, L.H., van Heel, D.A., Wijmenga, C., Meerman, G.J. Te, Franke, L., 2011. Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. PLoS Genet. 7, e1002197. doi:10.1371/journal.pgen.1002197

Fellay, J., Shianna, K.V., Ge, D., Colombo, S., Ledergerber, B., Weale, M., Zhang, K., Gumbs, C., Castagna, A., Cossarizza, A., Cozzi-Lepri, A., De Luca, A., Easterbrook, P., Francioli, P., Mallal, S., Martinez-Picado, J., Miro, J.M., Obel, N., Smith, J.P., Wyniger, J., Descombes, P., Antonarakis, S.E., Letvin, N.L., McMichael, A.J., Haynes, B.F., Telenti, A., Goldstein, D.B., 2007. A whole-genome association study of major determinants for host control of HIV-1. Science 317, 944–947. doi:10.1126/science.1143767

Filipe-Santos, O., Bustamante, J., Haverkamp, M.H., Vinolo, E., Ku, C.-L., Puel, A., Frucht, D.M., Christel, K., von Bernuth, H., Jouanguy, E., Feinberg, J., Durandy, A., Senechal, B., Chapgier, A., Vogt, G., de Beaucoudrey, L., Fieschi, C., Picard, C., Garfa, M., Chemli, J., Bejaoui, M., Tsolia, M.N., Kutukculer, N., Plebani, A., Notarangelo, L., Bodemer, C., Geissmann, F., Israël, A., Véron, M., Knackstedt, M., Barbouche, R., Abel, L., Magdorf, K., Gendrel, D., Agou, F., Holland, S.M., Casanova, J.-L., 2006. X-linked susceptibility to mycobacteria is caused by mutations in NEMO impairing CD40-dependent IL-12 production. J. Exp. Med. 203, 1745–1759. doi:10.1084/jem.20060085

Filipowicz, W., Bhattacharyya, S.N., Sonenberg, N., 2008. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat. Rev. Genet. 9, 102–114. doi:10.1038/nrg2290

Fischer, A., Wiebe, V., Pääbo, S., Przeworski, M., 2004. Evidence for a complex demographic history of chimpanzees. Mol. Biol. Evol. 21, 799–808. doi:10.1093/molbev/msh083

Flores-Villanueva, P.O., Ruiz-Morales, J.A., Song, C.-H., Flores, L.M., Jo, E.-K., Montaño, M., Barnes, P.F., Selman, M., Granados, J., 2005. A functional promoter polymorphism in monocyte chemoattractant protein–1 is associated with increased susceptibility to pulmonary tuberculosis. J Exp Med 202, 1649–1658. doi:10.1084/jem.20050126

Flory, C.M., Hubbard, R.D., Collins, F.M., 1992. Effects of in vivo T lymphocyte subset depletion on mycobacterial infections in mice. J. Leukoc. Biol. 51, 225–229.

Fornarino, S., Laval, G., Barreiro, L.B., Manry, J., Vasseur, E., Quintana-Murci, L., 2011. Evolution of the TIR domain-containing adaptors in humans: swinging between

constraint and adaptation. Mol. Biol. Evol. 28, 3087–3097. doi:10.1093/molbev/msr137

Fortin, A., Abel, L., Casanova, J.L., Gros, P., 2007. Host genetics of mycobacterial diseases in mice and men: forward genetic studies of BCG-osis and tuberculosis. Annu Rev Genomics Hum Genet 8, 163–192. doi:10.1146/annurev.genom.8.080706.092315

Fraser, H.B., 2013. Gene expression drives local adaptation in humans. Genome Res. 23, 1089–1096. doi:10.1101/gr.152710.112

Friedberg, E.C., 2003. DNA damage and repair. Nature 421, 436–440. doi:10.1038/nature01408

Friedberg, E.C., Walker, C.G., Siede, W., Wood, R.D., Schultz, R.A., Ellenburger, T., 2006. DNA repair and mutagenesis, 2nd ed. ed. ASM Press, Washington, D.C.

Fry, A.E., Ghansa, A., Small, K.S., Palma, A., Auburn, S., Diakite, M., Green, A., Campino, S., Teo, Y.Y., Clark, T.G., Jeffreys, A.E., Wilson, J., Jallow, M., Sisay-Joof, F., Pinder, M., Griffiths, M.J., Peshu, N., Williams, T.N., Newton, C.R., Marsh, K., Molyneux, M.E., Taylor, T.E., Koram, K.A., Oduro, A.R., Rogers, W.O., Rockett, K.A., Sabeti, P.C., Kwiatkowski, D.P., 2009. Positive selection of a CD36 nonsense variant in sub-Saharan Africa, but no association with severe malaria phenotypes. Human molecular genetics 18, 2683–92. doi:10.1093/hmg/ddp192

Fumagalli, M., Cagliani, R., Pozzoli, U., Riva, S., Comi, G.P., Menozzi, G., Bresolin, N., Sironi, M., 2009a. Widespread balancing selection and pathogen-driven selection at blood group antigen genes. Genome Res. 19, 199–212. doi:10.1101/gr.082768.108

Fumagalli, M., Cagliani, R., Riva, S., Pozzoli, U., Biasin, M., Piacentini, L., Comi, G.P., Bresolin, N., Clerici, M., Sironi, M., 2010. Population genetics of IFIH1: ancient population structure, local selection, and implications for susceptibility to type 1 diabetes. Mol. Biol. Evol. 27, 2555–2566. doi:10.1093/molbev/msq141

Fumagalli, M., Pozzoli, U., Cagliani, R., Comi, G.P., Riva, S., Clerici, M., Bresolin, N., Sironi, M., 2009b. Parasites represent a major selective force for interleukin genes and shape the genetic predisposition to autoimmune conditions. J. Exp. Med. 206, 1395–1408. doi:10.1084/jem.20082779

Fumagalli, M., Sironi, M., Pozzoli, U., Ferrer-Admetlla, A., Ferrer-Admettla, A., Pattini, L., Nielsen, R., 2011. Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. PLoS Genet. 7, e1002355. doi:10.1371/journal.pgen.1002355

Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S.M., Bondarev, A.A., Johnson, P.L.F., Aximu-Petri, A., Prüfer, K., de Filippo, C., Meyer, M., Zwyns, N., Salazar-García, D.C., Kuzmin, Y.V., Keates, S.G., Kosintsev, P.A., Razhev, D.I., Richards, M.P., Peristov, N.V., Lachmann, M., Douka, K., Higham, T.F.G., Slatkin, M., Hublin, J.-J., Reich, D., Kelso, J., Viola, T.B., Pääbo, S., 2014. Genome sequence of a 45,000-year-old modern human from western Siberia. Nature 514, 445–449. doi:10.1038/nature13810

Fu, W., Akey, J.M., 2013. Selection and Adaptation in the Human Genome. Annual Review of Genomics and Human Genetics 14, 467–489. doi:10.1146/annurev-genom-091212-153509

Fu, Y.X., Li, W.H., 1993. Statistical tests of neutrality of mutations. Genetics 133, 693–709.

Fu, Y., Yi, Z., Wu, X., Li, J., Xu, F., 2011. Circulating microRNAs in patients with active pulmonary tuberculosis. J. Clin. Microbiol. 49, 4246–4251. doi:10.1128/JCM.05459-11

Gaffney, D.J., 2013. Global properties and functional complexity of human gene regulatory variation. PLoS Genet. 9, e1003501. doi:10.1371/journal.pgen.1003501

Ganz, T., Lehrer, R.I., 1994. Defensins. Curr. Opin. Immunol. 6, 584–589.

Gargalovic, P.S., Imura, M., Zhang, B., Gharavi, N.M., Clark, M.J., Pagnon, J., Yang, W.-P., He, A., Truong, A., Patel, S., Nelson, S.F., Horvath, S., Berliner, J.A., Kirchgessner, T.G., Lusis, A.J., 2006. Identification of inflammatory gene modules based on variations of human endothelial cell responses to oxidized lipids. Proc. Natl. Acad. Sci. U.S.A. 103, 12741–12746. doi:10.1073/pnas.0605457103

Garud, N.R., Messer, P.W., Buzbas, E.O., Petrov, D.A., 2015. Recent selective sweeps in North American Drosophila melanogaster show signatures of soft sweeps. PLoS Genet. 11, e1005004. doi:10.1371/journal.pgen.1005004

Garud, N.R., Rosenberg, N.A., 2015. Enhancing the mathematical properties of new haplotype homozygosity statistics for the detection of selective sweeps. Theor Popul Biol 102, 94–101. doi:10.1016/j.tpb.2015.04.001

Ge, D., Fellay, J., Thompson, A.J., Simon, J.S., Shianna, K.V., Urban, T.J., Heinzen, E.L., Qiu, P., Bertelsen, A.H., Muir, A.J., Sulkowski, M., McHutchison, J.G., Goldstein, D.B., 2009. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. Nature 461, 399–401. doi:10.1038/nature08309

Giacomini, E., Sotolongo, A., Iona, E., Severa, M., Remoli, M.E., Gafa, V., Lande, R., Fattorini, L., Smith, I., Manganelli, R., Coccia, E.M., 2006. Infection of human dendritic cells with a Mycobacterium tuberculosis sigE mutant stimulates production of high levels of interleukin-10 but low levels of CXCL10: impact on the T-cell response. Infect. Immun. 74, 3296–3304. doi:10.1128/IAI.01687-05

Gilad, Y., Rifkin, S.A., Pritchard, J.K., 2008. Revealing the architecture of gene regulation: the promise of eQTL studies. Trends Genet. 24, 408–415. doi:10.1016/j.tig.2008.06.001

Girardin, S.E., Tournebize, R., Mavris, M., Page, A.L., Li, X., Stark, G.R., Bertin, J., DiStefano, P.S., Yaniv, M., Sansonetti, P.J., Philpott, D.J., 2001. CARD4/Nod1 mediates NF-kappaB and JNK activation by invasive Shigella flexneri. EMBO Rep. 2, 736–742. doi:10.1093/embo-reports/kve155

Godfroy, J.I., Roostan, M., Moroz, Y.S., Korendovych, I.V., Yin, H., 2012. Isolated Toll-like receptor transmembrane domains are capable of oligomerization. PLoS ONE 7, e48875. doi:10.1371/journal.pone.0048875

Grad, Y., Aach, J., Hayes, G.D., Reinhart, B.J., Church, G.M., Ruvkun, G., Kim, J., 2003. Computational and experimental identification of C. elegans microRNAs. Mol. Cell 11, 1253–1263.

Graham, L.M., Brown, G.D., 2009. The Dectin-2 family of C-type lectins in immunity and homeostasis. Cytokine 48, 148–155. doi:10.1016/j.cyto.2009.07.010

Grant, A.V., Baghdadi, J. El, Sabri, A., Azbaoui, S. El, Alaoui-Tahiri, K., Abderrahmani Rhorfi, I., Gharbaoui, Y., Abid, A., Benkirane, M., Raharimanga, V., Richard, V., Orlova, M., Boland, A., Migaud, M., Okada, S., Nolan, D.K., Bustamante, J.,

Barreiro, L.B., Schurr, E., Boisson-Dupuis, S., Rasolofo, V., Casanova, J.L., Abel, L., 2013. Age-dependent association between pulmonary tuberculosis and common TOX variants in the 8q12-13 linkage region. American journal of human genetics 92, 407–14. doi:10.1016/j.ajhg.2013.01.013

Gravel, S., Henn, B.M., Gutenkunst, R.N., Indap, A.R., Marth, G.T., Clark, A.G., Yu, F., Gibbs, R.A., 1000 Genomes Project, Bustamante, C.D., 2011. Demographic history and rare allele sharing among human populations. Proc. Natl. Acad. Sci. U.S.A. 108, 11983–11988. doi:10.1073/pnas.1019276108

Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H.-Y., Hansen, N.F., Durand, E.Y., Malaspinas, A.-S., Jensen, J.D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H.A., Good, J.M., Schultz, R., Aximu-Petri, A., Butthof, A., Höber, B., Höffner, B., Siegemund, M., Weihmann, A., Nusbaum, C., Lander, E.S., Russ, C., Novod, N., Affourtit, J., Egholm, M., Verna, C., Rudan, P., Brajkovic, D., Kucan, Z., Gusic, I., Doronichev, V.B., Golovanova, L.V., Lalueza-Fox, C., de la Rasilla, M., Fortea, J., Rosas, A., Schmitz, R.W., Johnson, P.L.F., Eichler, E.E., Falush, D., Birney, E., Mullikin, J.C., Slatkin, M., Nielsen, R., Kelso, J., Lachmann, M., Reich, D., Pääbo, S., 2010. A draft sequence of the Neandertal genome. Science 328, 710–722. doi:10.1126/science.1188021

Griffiths-Jones, S., Saini, H.K., van Dongen, S., Enright, A.J., 2008. miRBase: tools for microRNA genomics. Nucleic Acids Res. 36, D154–158. doi:10.1093/nar/gkm952

Grimson, A., Farh, K.K.-H., Johnston, W.K., Garrett-Engele, P., Lim, L.P., Bartel, D.P., 2007. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol. Cell 27, 91–105. doi:10.1016/j.molcel.2007.06.017

Gringhuis, S.I., Kaptein, T.M., Wevers, B.A., Theelen, B., van der Vlist, M., Boekhout, T., Geijtenbeek, T.B.H., 2012. Dectin-1 is an extracellular pathogen sensor for the induction and processing of IL-1β via a noncanonical caspase-8 inflammasome. Nat. Immunol. 13, 246–254. doi:10.1038/ni.2222

Grossman, S.R., Andersen, K.G., Shlyakhter, I., Tabrizi, S., Winnicki, S., Yen, A., Park, D.J., Griesemer, D., Karlsson, E.K., Wong, S.H., Cabili, M., Adegbola, R.A., Bamezai, R.N., Hill, A.V., Vannberg, F.O., Rinn, J.L., Lander, E.S., Schaffner, S.F., Sabeti, P.C., 2013. Identifying recent adaptations in large-scale genomic data. Cell 152, 703–13. doi:10.1016/j.cell.2013.01.035

Grossman, S.R., Shlyakhter, I., Karlsson, E.K., Byrne, E.H., Morales, S., Frieden, G., Hostetter, E., Angelino, E., Garber, M., Zuk, O., Lander, E.S., Schaffner, S.F., Sabeti, P.C., 2010. A composite of multiple signals distinguishes causal variants in regions of positive selection. Science 327, 883–6. doi:10.1126/science.1183863

Guan, Y., Ranoa, D.R.E., Jiang, S., Mutha, S.K., Li, X., Baudry, J., Tapping, R.I., 2010. Human TLRs 10 and 1 Share Common Mechanisms of Innate Immune Sensing but Not Signaling. J Immunol 184, 5094–5103. doi:10.4049/jimmunol.0901888

Guduric-Fuchs, J., O'Connor, A., Camp, B., O'Neill, C.L., Medina, R.J., Simpson, D.A., 2012. Selective extracellular vesicle-mediated export of an overlapping set of microRNAs from multiple cell types. BMC Genomics 13, 357. doi:10.1186/1471-2164-13-357

Gunnarsdóttir, E.D., Li, M., Bauchet, M., Finstermeier, K., Stoneking, M., 2011. High-throughput sequencing of complete human mtDNA genomes from the Philippines. Genome Res. 21, 1–11. doi:10.1101/gr.107615.110

Haas, F., Haas, S.S., 1996. The origins of Mycobacterium tuberculosis and the notion of its contagiousness 3–19.

Haldane, J.B.S., 1949. Disease and evolution. Ric. Sci. 19, 68–76.

Hambleton, S., Salem, S., Bustamante, J., Bigley, V., Boisson-Dupuis, S., Azevedo, J., Fortin, A., Haniffa, M., Ceron-Gutierrez, L., Bacon, C.M., Menon, G., Trouillet, C., McDonald, D., Carey, P., Ginhoux, F., Alsina, L., Zumwalt, T.J., Kong, X.-F., Kumararatne, D., Butler, K., Hubeau, M., Feinberg, J., Al-Muhsen, S., Cant, A., Abel, L., Chaussabel, D., Doffinger, R., Talesnik, E., Grumach, A., Duarte, A., Abarca, K., Moraes-Vasconcelos, D., Burk, D., Berghuis, A., Geissmann, F., Collin, M., Casanova, J.-L., Gros, P., 2011. IRF8 mutations and human dendritic-cell immunodeficiency. N. Engl. J. Med. 365, 127–138. doi:10.1056/NEJMoa1100066

Hamblin, M.T., Di Rienzo, A., 2000. Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. American journal of human genetics 66, 1669–79. doi:10.1086/302879

Hammer, M.F., Woerner, A.E., Mendez, F.L., Watkins, J.C., Wall, J.D., 2011. Genetic evidence for archaic admixture in Africa. Proc. Natl. Acad. Sci. U.S.A. 108, 15123–15128. doi:10.1073/pnas.1109300108

Han, J., Lee, Y., Yeom, K.-H., Nam, J.-W., Heo, I., Rhee, J.-K., Sohn, S.Y., Cho, Y., Zhang, B.-T., Kim, V.N., 2006. Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell 125, 887–901. doi:10.1016/j.cell.2006.03.043

Han, Y., Hazelett, D.J., Wiklund, F., Schumacher, F.R., Stram, D.O., Berndt, S.I., Wang, Z., Rand, K.A., Hoover, R.N., Machiela, M.J., Yeager, M., Burdette, L., Chung, C.C., Hutchinson, A., Yu, K., Xu, J., Travis, R.C., Key, T.J., Siddiq, A., Canzian, F., Takahashi, A., Kubo, M., Stanford, J.L., Kolb, S., Gapstur, S.M., Diver, W.R., Stevens, V.L., Strom, S.S., Pettaway, C.A., Olama, A.A. Al, Kote-Jarai, Z., Eeles, R.A., Yeboah, E.D., Tettey, Y., Biritwum, R.B., Adjei, A.A., Tay, E., Truelove, A., Niwa, S., Chokkalingam, A.P., Isaacs, W.B., Chen, C., Lindstrom, S., Le Marchand, L., Giovannucci, E.L., Pomerantz, M., Long, H., Li, F., Ma, J., Stampfer, M., John, E.M., Ingles, S.A., Kittles, R.A., Murphy, A.B., Blot, W.J., Signorello, L.B., Zheng, W., Albanes, D., Virtamo, J., Weinstein, S., Nemesure, B., Carpten, J., Leske, M.C., Wu, S.-Y., Hennis, A.J.M., Rybicki, B.A., Neslund-Dudas, C., Hsing, A.W., Chu, L., Goodman, P.J., Klein, E.A., Zheng, S.L., Witte, J.S., Casey, G., Riboli, E., Li, Q., Freedman, M.L., Hunter, D.J., Gronberg, H., Cook, M.B., Nakagawa, H., Kraft, P., Chanock, S.J., Easton, D.F., Henderson, B.E., Coetzee, G.A., Conti, D.V., Haiman, C.A., 2015. Integration of multiethnic fine-mapping and genomic annotation to prioritize candidate functional SNPs at prostate cancer susceptibility regions. Hum. Mol. Genet. doi:10.1093/hmg/ddv269

Hargreaves, J.R., Boccia, D., Evans, C.A., Adato, M., Petticrew, M., Porter, J.D.H., 2011. The Social Determinants of Tuberculosis: From Evidence to Action. American Journal of Public Health 101, 654–662. doi:10.2105/AJPH.2010.199505

Harpending, H., Rogers, A., 2000. Genetic perspectives on human origins and differentiation. Annu Rev Genomics Hum Genet 1, 361–385. doi:10.1146/annurev.genom.1.1.361

Hartford, C.M., Duan, S., Delaney, S.M., Mi, S., Kistner, E.O., Lamba, J.K., Huang, R.S., Dolan, M.E., 2009. Population-specific genetic variants important in susceptibility to cytarabine arabinoside cytotoxicity. Blood 113, 2145–2153. doi:10.1182/blood-2008-05-154302

Hasan, U., Chaffois, C., Gaillard, C., Saulnier, V., Merck, E., Tancredi, S., Guiet, C., Brière, F., Vlach, J., Lebecque, S., Trinchieri, G., Bates, E.E.M., 2005. Human TLR10 Is a Functional Receptor, Expressed by B Cells and Plasmacytoid Dendritic Cells, Which Activates Gene Transcription through MyD88. J Immunol 174, 2942–2950. doi:10.4049/jimmunol.174.5.2942

Hashimoto, C., Hudson, K.L., Anderson, K.V., 1988. The Toll gene of Drosophila, required for dorsal-ventral embryonic polarity, appears to encode a transmembrane protein. Cell 52, 269–279.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., Wang, W., Weng, Z., Green, R.D., Crawford, G.E., Ren, B., 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. 39, 311–318. doi:10.1038/ng1966

Hellenthal, G., Auton, A., Falush, D., 2008. Inferring human colonization history using a copying model. PLoS Genet. 4, e1000078. doi:10.1371/journal.pgen.1000078

Hermisson, J., Pennings, P.S., 2005. Soft Sweeps Molecular Population Genetics of Adaptation From Standing Genetic Variation. Genetics 169, 2335–2352. doi:10.1534/genetics.104.036947

Hernandez, R.D., Kelley, J.L., Elyashiv, E., Melton, S.C., Auton, A., McVean, G., 1000 Genomes Project, Sella, G., Przeworski, M., 2011. Classic selective sweeps were rare in recent human evolution. Science 331, 920–924. doi:10.1126/science.1198878

Hershkovitz, I., Donoghue, H.D., Minnikin, D.E., Besra, G.S., Lee, O.Y.-C., Gernaey, A.M., Galili, E., Eshed, V., Greenblatt, C.L., Lemma, E., Bar-Gal, G.K., Spigelman, M., 2008. Detection and Molecular Characterization of 9000-Year-Old Mycobacterium tuberculosis from a Neolithic Settlement in the Eastern Mediterranean. PLoS ONE 3, e3426. doi:10.1371/journal.pone.0003426

He, X., Jing, Z., Cheng, G., 2014. MicroRNAs: new regulators of Toll-like receptor signalling pathways. Biomed Res Int 2014, 945169. doi:10.1155/2014/945169

Hibberd, M.L., Sumiya, M., Summerfield, J.A., Booy, R., Levin, M., 1999. Association of variants of the gene for mannose-binding lectin with susceptibility to meningococcal disease. Meningococcal Research Group. Lancet 353, 1049–1053.

Higham, T., Douka, K., Wood, R., Ramsey, C.B., Brock, F., Basell, L., Camps, M., Arrizabalaga, A., Baena, J., Barroso-Ruíz, C., Bergman, C., Boitard, C., Boscato, P., Caparrós, M., Conard, N.J., Draily, C., Froment, A., Galván, B., Gambassini, P., Garcia-Moreno, A., Grimaldi, S., Haesaerts, P., Holt, B., Iriarte-Chiapusso, M.-J., Jelinek, A., Jordá Pardo, J.F., Maíllo-Fernández, J.-M., Marom, A., Maroto, J., Menéndez, M., Metz, L., Morin, E., Moroni, A., Negrino, F., Panagopoulou, E., Peresani, M., Pirson, S., de la Rasilla, M., Riel-Salvatore, J., Ronchitelli, A., Santamaria, D., Semal, P., Slimak, L., Soler, J., Soler, N., Villaluenga, A., Pinhasi, R., Jacobi, R., 2014. The timing and spatiotemporal patterning of Neanderthal disappearance. Nature 512, 306–309. doi:10.1038/nature13621

Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S., Manolio, T.A., 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc. Natl. Acad. Sci. U.S.A. 106, 9362–9367. doi:10.1073/pnas.0903103106

Hiscott, J., Lacoste, J., Lin, R., 2006. Recruitment of an interferon molecular signaling complex to the mitochondrial membrane: disruption by hepatitis C virus NS3-4A protease. Biochem. Pharmacol. 72, 1477–1484. doi:10.1016/j.bcp.2006.06.030

Hoal-Van Helden, E.G., Epstein, J., Victor, T.C., Hon, D., Lewis, L.A., Beyers, N., Zurakowski, D., Ezekowitz, A.B., Van Helden, P.D., 1999. Mannose-binding protein B allele confers protection against tuberculous meningitis. Pediatr. Res. 45, 459–464. doi:10.1203/00006450-199904010-00002

Hoeppner, V.H., Marciniuk, D.D., 2000. Tuberculosis in aboriginal Canadians. Can. Respir. J. 7, 141–146.

Hoffmann, J.A., 2003. The immune response of Drosophila. Nature 426, 33–38. doi:10.1038/nature02021

Hoffmann, J.A., Kafatos, F.C., Janeway, C.A., Ezekowitz, R.A., 1999. Phylogenetic perspectives in innate immunity. Science 284, 1313–1318.

Hollox, E.J., Armour, J.A.L., 2008. Directional and balancing selection in human beta-defensins. BMC Evol. Biol. 8, 113. doi:10.1186/1471-2148-8-113

Horan, K.A., Hansen, K., Jakobsen, M.R., Holm, C.K., Søby, S., Unterholzner, L., Thompson, M., West, J.A., Iversen, M.B., Rasmussen, S.B., Ellermann-Eriksen, S., Kurt-Jones, E., Landolfo, S., Damania, B., Melchjorsen, J., Bowie, A.G., Fitzgerald, K.A., Paludan, S.R., 2013. Proteasomal degradation of herpes simplex virus capsids in macrophages releases DNA to the cytosol for recognition by DNA sensors. J. Immunol. 190, 2311–2319. doi:10.4049/jimmunol.1202749

Hornung, V., Ablasser, A., Charrel-Dennis, M., Bauernfeind, F., Horvath, G., Caffrey, D.R., Latz, E., Fitzgerald, K.A., 2009. AIM2 recognizes cytosolic dsDNA and forms a caspase-1-activating inflammasome with ASC. Nature 458, 514–518. doi:10.1038/nature07725

Hornung, V., Rothenfusser, S., Britsch, S., Krug, A., Jahrsdörfer, B., Giese, T., Endres, S., Hartmann, G., 2002. Quantitative Expression of Toll-Like Receptor 1–10 mRNA in Cellular Subsets of Human Peripheral Blood Mononuclear Cells and Sensitivity to CpG Oligodeoxynucleotides. J Immunol 168, 4531–4537. doi:10.4049/jimmunol.168.9.4531

Hsu, Y.-M.S., Zhang, Y., You, Y., Wang, D., Li, H., Duramad, O., Qin, X.-F., Dong, C., Lin, X., 2007. The adaptor protein CARD9 is required for innate immune responses to intracellular pathogens. Nat. Immunol. 8, 198–205. doi:10.1038/ni1426

Huerta-Sánchez, E., Jin, X., Asan, null, Bianba, Z., Peter, B.M., Vinckenbosch, N., Liang, Y., Yi, X., He, M., Somel, M., Ni, P., Wang, B., Ou, X., Huasang, null, Luosang, J., Cuo, Z.X.P., Li, K., Gao, G., Yin, Y., Wang, W., Zhang, X., Xu, X., Yang, H., Li, Y., Wang, J., Wang, J., Nielsen, R., 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature 512, 194–197. doi:10.1038/nature13408

Hulur, I., Gamazon, E.R., Skol, A.D., Xicola, R.M., Llor, X., Onel, K., Ellis, N.A., Kupfer, S.S., 2015. Enrichment of inflammatory bowel disease and colorectal cancer risk variants in colon expression quantitative trait loci. BMC Genomics 16, 138. doi:10.1186/s12864-015-1292-z

Huntzinger, E., Izaurralde, E., 2011. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nat. Rev. Genet. 12, 99–110. doi:10.1038/nrg2936

Hutvágner, G., McLachlan, J., Pasquinelli, A.E., Bálint, E., Tuschl, T., Zamore, P.D., 2001. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science 293, 834–838. doi:10.1126/science.1062961

Huysamen, C., Brown, G.D., 2009. The fungal pattern recognition receptor, Dectin-1, and the associated cluster of C-type lectin-like receptors. FEMS Microbiol. Lett. 290, 121–128. doi:10.1111/j.1574-6968.2008.01418.x

Idaghdour, Y., Awadalla, P., 2012. Exploiting gene expression variation to capture gene-environment interactions for disease. Front Genet 3, 228. doi:10.3389/fgene.2012.00228

Idaghdour, Y., Quinlan, J., Goulet, J.-P., Berghout, J., Gbeha, E., Bruat, V., de Malliard, T., Grenier, J.-C., Gomez, S., Gros, P., Rahimy, M.C., Sanni, A., Awadalla, P., 2012. Evidence for additive and interaction effects of host genotype and infection in malaria. Proc. Natl. Acad. Sci. U.S.A. 109, 16786–16793. doi:10.1073/pnas.1204945109

Ikemura, T., 1981. Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the E. coli translational system. J. Mol. Biol. 151, 389–409.

Ingman, M., Gyllensten, U., 2001. Analysis of the complete human mtDNA genome: methodology and inferences for human evolution. J. Hered. 92, 454–461.

Ingman, M., Kaessmann, H., Pääbo, S., Gyllensten, U., 2000. Mitochondrial genome variation and the origin of modern humans. Nature 408, 708–713. doi:10.1038/35047064

Innan, H., Kim, Y., 2008. Detecting local adaptation using the joint sampling of polymorphism data in the parental and derived populations. Genetics 179, 1713–1720. doi:10.1534/genetics.108.086835

Inohara, N., Koseki, T., Lin, J., del Peso, L., Lucas, P.C., Chen, F.F., Ogura, Y., Núñez, G., 2000. An induced proximity model for NF-kappa B activation in the Nod1/RICK and RIP signaling pathways. J. Biol. Chem. 275, 27823–27831. doi:10.1074/jbc.M003415200

International HapMap 3 Consortium, 2010. Integrating common and rare genetic variation in diverse human populations. Nature 467, 52–58. doi:10.1038/nature09298

International HapMap Consortium, 2007. A second generation human haplotype map of over 3.1 million SNPs. Nature 449, 851–861. doi:10.1038/nature06258

International HapMap Consortium, 2005. A haplotype map of the human genome. Nature 437, 1299–1320. doi:10.1038/nature04226

Ishikawa, H., Barber, G.N., 2008. STING is an endoplasmic reticulum adaptor that facilitates innate immune signalling. Nature 455, 674–678. doi:10.1038/nature07317

Ishikawa, H., Ma, Z., Barber, G.N., 2009. STING regulates intracellular DNA-mediated, type I interferon-dependent innate immunity. Nature 461, 788–792. doi:10.1038/nature08476

Jallow, M., Teo, Y.Y., Small, K.S., Rockett, K.A., Deloukas, P., Clark, T.G., Kivinen, K., Bojang, K.A., Conway, D.J., Pinder, M., Sirugo, G., Sisay-Joof, F., Usen, S., Auburn, S., Bumpstead, S.J., Campino, S., Coffey, A., Dunham, A., Fry, A.E., Green, A., Gwilliam, R., Hunt, S.E., Inouye, M., Jeffreys, A.E., Mendy, A., Palotie, A., Potter, S., Ragoussis, J., Rogers, J., Rowlands, K., Somaskantharajah, E., Whittaker, P., Widden, C., Donnelly, P., Howie, B., Marchini, J., Morris, A., SanJoaquin, M., Achidi, E.A., Agbenyega, T., Allen, A., Amodu, O., Corran, P., Djimde, A., Dolo, A., Doumbo, O.K., Drakeley, C., Dunstan, S., Evans, J., Farrar, J., Fernando, D., Hien, T.T., Horstmann, R.D., Ibrahim, M., Karunaweera, N., Kokwaro, G., Koram, K.A., Lemnge, M., Makani, J., Marsh, K., Michon, P., Modiano, D., Molyneux, M.E., Mueller, I., Parker, M., Peshu, N., Plowe, C.V., Puijalon, O., Reeder, J., Reyburn, H., Riley, E.M., Sakuntabhai, A., Singhasivanon, P., Sirima, S., Tall, A., Taylor, T.E., Thera, M., Troye-Blomberg, M., Williams, T.N., Wilson, M., Kwiatkowski, D.P., Wellcome Trust Case Control Consortium, Malaria Genomic Epidemiology Network, 2009. Genome-wide and fine-resolution association analysis of malaria in West Africa. Nat. Genet. 41, 657–665. doi:10.1038/ng.388

Janeway, C.A., 2001. How the immune system protects the host from infection. Microbes and Infection 3, 1167–1171. doi:10.1016/S1286-4579(01)01477-0

Janeway, C.A., 1989. Approaching the asymptote? Evolution and revolution in immunology. Cold Spring Harb. Symp. Quant. Biol. 54 Pt 1, 1–13.

Janeway, C.A., Medzhitov, R., 2002. Innate immune recognition. Annu. Rev. Immunol. 20, 197–216. doi:10.1146/annurev.immunol.20.083001.084359

Jansen, R.C., Nap, J.P., 2001. Genetical genomics: the added value from segregation. Trends Genet. 17, 388–391.

Jobling, M.A., Tyler-Smith, C., 2003. The human Y chromosome: an evolutionary marker comes of age. Nat. Rev. Genet. 4, 598–612. doi:10.1038/nrg1124

Jobling, M., Hurles, M., Tyler-Smith, C., 2013. Human Evolutionary Genetics: Origins, Peoples & Disease. Garland Science.

Johnnidis, J.B., Harris, M.H., Wheeler, R.T., Stehling-Sun, S., Lam, M.H., Kirak, O., Brummelkamp, T.R., Fleming, M.D., Camargo, F.D., 2008. Regulation of progenitor cell proliferation and granulocyte function by microRNA-223. Nature 451, 1125–1129. doi:10.1038/nature06607

Johnson, C.M., Lyle, E.A., Omueti, K.O., Stepensky, V.A., Yegin, O., Alpsoy, E., Hamann, L., Schumann, R.R., Tapping, R.I., 2007. Cutting edge: A common polymorphism impairs cell surface trafficking and functional responses of TLR1 but protects against leprosy. J. Immunol. 178, 7520–7524.

Johnson, D.S., Mortazavi, A., Myers, R.M., Wold, B., 2007. Genome-wide mapping of in vivo protein-DNA interactions. Science 316, 1497–1502. doi:10.1126/science.1141319

Jopling, C., 2012. Liver-specific microRNA-122: Biogenesis and function. RNA Biol 9, 137–142. doi:10.4161/rna.18827

Jopling, C.L., Yi, M., Lancaster, A.M., Lemon, S.M., Sarnow, P., 2005. Modulation of Hepatitis C Virus RNA Abundance by a Liver-Specific MicroRNA. Science 309, 1577–1581. doi:10.1126/science.1113329

Jouanguy, E., Altare, F., Lamhamedi, S., Revy, P., Emile, J.F., Newport, M., Levin, M., Blanche, S., Seboun, E., Fischer, A., Casanova, J.L., 1996. Interferon-gamma-receptor deficiency in an infant with fatal bacille Calmette-Guérin infection. N. Engl. J. Med. 335, 1956–1961. doi:10.1056/NEJM199612263352604

Kanazawa, N., 2007. Dendritic cell immunoreceptors: C-type lectin receptors for pattern-recognition and signaling on antigen-presenting cells. J. Dermatol. Sci. 45, 77–86. doi:10.1016/j.jdermsci.2006.09.001

Kanazawa, N., Tashiro, K., Inaba, K., Miyachi, Y., 2003. Dendritic cell immunoactivating receptor, a novel C-type lectin immunoreceptor, acts as an activating receptor through association with Fc receptor gamma chain. J. Biol. Chem. 278, 32645–32652. doi:10.1074/jbc.M304226200

Kanazawa, N., Tashiro, K., Miyachi, Y., 2004. Signaling and immune regulatory role of the dendritic cell immunoreceptor (DCIR) family lectins: DCIR, DCAR, dectin-2 and BDCA-2. Immunobiology 209, 179–190. doi:10.1016/j.imbio.2004.03.004

Kang, D., Gopalkrishnan, R.V., Wu, Q., Jankowsky, E., Pyle, A.M., Fisher, P.B., 2002. mda-5: An interferon-inducible putative RNA helicase with double-stranded RNA-dependent ATPase activity and melanoma growth-suppressive properties. Proc. Natl. Acad. Sci. U.S.A. 99, 637–642. doi:10.1073/pnas.022637199

Karikó, K., Buckstein, M., Ni, H., Weissman, D., 2005. Suppression of RNA recognition by Toll-like receptors: the impact of nucleoside modification and the evolutionary origin of RNA. Immunity 23, 165–175. doi:10.1016/j.immuni.2005.06.008

Kato, H., Sato, S., Yoneyama, M., Yamamoto, M., Uematsu, S., Matsui, K., Tsujimura, T., Takeda, K., Fujita, T., Takeuchi, O., Akira, S., 2005. Cell type-specific involvement of RIG-I in antiviral response. Immunity 23, 19–28. doi:10.1016/j.immuni.2005.04.010

Kawai, T., Akira, S., 2011. Toll-like receptors and their crosstalk with other innate receptors in infection and immunity. Immunity 34, 637–650. doi:10.1016/j.immuni.2011.05.006

Kawai, T., Akira, S., 2010. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. Nat. Immunol. 11, 373–384. doi:10.1038/ni.1863

Kawai, T., Akira, S., 2006. Innate immune recognition of viral infection. Nat. Immunol. 7, 131–137. doi:10.1038/ni1303

Kawai, T., Takahashi, K., Sato, S., Coban, C., Kumar, H., Kato, H., Ishii, K.J., Takeuchi, O., Akira, S., 2005. IPS-1, an adaptor triggering RIG-I- and Mda5-mediated type I interferon induction. Nat. Immunol. 6, 981–988. doi:10.1038/ni1243

Keating, S.E., Baran, M., Bowie, A.G., 2011. Cytosolic DNA sensors regulating type I interferon induction. Trends Immunol. 32, 574–581. doi:10.1016/j.it.2011.08.004

Kellis, M., Wold, B., Snyder, M.P., Bernstein, B.E., Kundaje, A., Marinov, G.K., Ward, L.D., Birney, E., Crawford, G.E., Dekker, J., Dunham, I., Elnitski, L.L., Farnham,

P.J., Feingold, E.A., Gerstein, M., Giddings, M.C., Gilbert, D.M., Gingeras, T.R., Green, E.D., Guigo, R., Hubbard, T., Kent, J., Lieb, J.D., Myers, R.M., Pazin, M.J., Ren, B., Stamatoyannopoulos, J.A., Weng, Z., White, K.P., Hardison, R.C., 2014. Defining functional DNA elements in the human genome. Proceedings of the National Academy of Sciences 111, 6131–6138. doi:10.1073/pnas.1318948111

Kerrigan, A.M., Brown, G.D., 2011. Syk-coupled C-type lectins in immunity. Trends Immunol. 32, 151–156. doi:10.1016/j.it.2011.01.002

Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J., Plasterk, R.H., 2001. Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. Genes Dev. 15, 2654–2659. doi:10.1101/gad.927801

Khoury, M.J., Beaty, T.H., Cohen, B.H., 1993. Fundamentals of genetic epidemiology. Oxford University Press.

Khrameeva, E.E., Bozek, K., He, L., Yan, Z., Jiang, X., Wei, Y., Tang, K., Gelfand, M.S., Prufer, K., Kelso, J., Paabo, S., Giavalisco, P., Lachmann, M., Khaitovich, P., 2014. Neanderthal ancestry drives evolution of lipid catabolism in contemporary Europeans. Nat Commun 5. doi:10.1038/ncomms4584

Khvorova, A., Reynolds, A., Jayasena, S.D., 2003. Functional siRNAs and miRNAs exhibit strand bias. Cell 115, 209–216.

Kim, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D., Ren, B., 2005. A high-resolution map of active promoters in the human genome. Nature 436, 876–880. doi:10.1038/nature03877

Kim, T., Pazhoor, S., Bao, M., Zhang, Z., Hanabuchi, S., Facchinetti, V., Bover, L., Plumas, J., Chaperot, L., Qin, J., Liu, Y.-J., 2010. Aspartate-glutamate-alanine-histidine box motif (DEAH)/RNA helicase A helicases sense microbial DNA in human plasmacytoid dendritic cells. Proc. Natl. Acad. Sci. U.S.A. 107, 15181–15186. doi:10.1073/pnas.1006539107

Kimura, M., 1968a. Evolutionary rate at the molecular level. Nature 217, 624–626.

Kimura, M., 1968b. Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. Genet. Res. 11, 247–269.

King, M.C., Wilson, A.C., 1975. Evolution at two levels in humans and chimpanzees. Science 188, 107–116.

Kobayashi, K., Inohara, N., Hernandez, L.D., Galán, J.E., Núñez, G., Janeway, C.A., Medzhitov, R., Flavell, R.A., 2002. RICK/Rip2/CARDIAK mediates signalling for receptors of the innate and adaptive immune systems. Nature 416, 194–199. doi:10.1038/416194a

Koch, R., 1882. Die Aetiologie der Tuberkulose. Berl Klin Wochenschr 11, 221–230.

Kohn, K.W., Spears, C.L., Doty, P., 1966. Inter-strand crosslinking of DNA by nitrogen mustard. J. Mol. Biol. 19, 266–288.

Kosiol, C., Vinar, T., da Fonseca, R.R., Hubisz, M.J., Bustamante, C.D., Nielsen, R., Siepel, A., 2008. Patterns of positive selection in six Mammalian genomes. PLoS Genet. 4, e1000144. doi:10.1371/journal.pgen.1000144

Kovacsovics, M., Martinon, F., Micheau, O., Bodmer, J.L., Hofmann, K., Tschopp, J., 2002. Overexpression of Helicard, a CARD-containing helicase cleaved during apoptosis, accelerates DNA degradation. Curr. Biol. 12, 838–843.

Krieg, A.M., Vollmer, J., 2007. Toll-like receptors 7, 8, and 9: linking innate immunity to autoimmunity. Immunol. Rev. 220, 251–269. doi:10.1111/j.1600-065X.2007.00572.x

Krützfeldt, J., Rajewsky, N., Braich, R., Rajeev, K.G., Tuschl, T., Manoharan, M., Stoffel, M., 2005. Silencing of microRNAs in vivo with "antagomirs." Nature 438, 685–689. doi:10.1038/nature04303

Kudaravalli, S., Veyrieras, J.-B., Stranger, B.E., Dermitzakis, E.T., Pritchard, J.K., 2009. Gene expression levels are a target of recent natural selection in the human genome. Mol. Biol. Evol. 26, 649–658. doi:10.1093/molbev/msn289

Kufer, T.A., Sansonetti, P.J., 2011. NLR functions beyond pathogen recognition. Nat. Immunol. 12, 121–128. doi:10.1038/ni.1985

Kumar, H., Kawai, T., Akira, S., 2011. Pathogen recognition by the innate immune system. Int. Rev. Immunol. 30, 16–34. doi:10.3109/08830185.2010.529976

Kwan, T., Grundberg, E., Koka, V., Ge, B., Lam, K.C.L., Dias, C., Kindmark, A., Mallmin, H., Ljunggren, O., Rivadeneira, F., Estrada, K., van Meurs, J.B., Uitterlinden, A., Karlsson, M., Ohlsson, C., Mellström, D., Nilsson, O., Pastinen, T., Majewski, J., 2009. Tissue effect on genetic control of transcript isoform variation. PLoS Genet. 5, e1000608. doi:10.1371/journal.pgen.1000608

Lachance, J., Vernot, B., Elbers, C.C., Ferwerda, B., Froment, A., Bodo, J.-M., Lema, G., Fu, W., Nyambo, T.B., Rebbeck, T.R., Zhang, K., Akey, J.M., Tishkoff, S.A., 2012. Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse African hunter-gatherers. Cell 150, 457–469. doi:10.1016/j.cell.2012.07.009

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., Tuschl, T., 2001. Identification of novel genes coding for small expressed RNAs. Science 294, 853–858. doi:10.1126/science.1064921

Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W., Tuschl, T., 2002. Identification of tissue-specific microRNAs from mouse. Curr. Biol. 12, 735–739.

Lai, E.C., Tomancak, P., Williams, R.W., Rubin, G.M., 2003. Computational identification of Drosophila microRNA genes. Genome Biol. 4, R42. doi:10.1186/gb-2003-4-7-r42

Lamason, R.L., Mohideen, M.A., Mest, J.R., Wong, A.C., Norton, H.L., Aros, M.C., Jurynec, M.J., Mao, X., Humphreville, V.R., Humbert, J.E., Sinha, S., Moore, J.L., Jagadeeswaran, P., Zhao, W., Ning, G., Makalowska, I., McKeigue, P.M., O'Donnell, D., Kittles, R., Parra, E.J., Mangini, N.J., Grunwald, D.J., Shriver, M.D., Canfield, V.A., Cheng, K.C., 2005. SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. Science 310, 1782–6. doi:10.1126/science.1116238

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S.,

Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de la Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Patrinos, A., Morgan, M.J., de Jong, P., Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y.J., Szustakowki, J., International Human Genome Sequencing Consortium, 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–921. doi:10.1038/35057062

Lande, R., Giacomini, E., Grassi, T., Remoli, M.E., Iona, E., Miettinen, M., Julkunen, I., Coccia, E.M., 2003. IFN-alpha beta released by Mycobacterium tuberculosis-infected human dendritic cells induces the expression of CXCL10: selective recruitment of NK and activated T cells. J. Immunol. 170, 1174–1182.

Lange, C., Hemmrich, G., Klostermeier, U.C., López-Quintero, J.A., Miller, D.J., Rahn, T., Weiss, Y., Bosch, T.C.G., Rosenstiel, P., 2011. Defining the origins of the NOD-like receptor system at the base of animal evolution. Mol. Biol. Evol. 28, 1687–1702. doi:10.1093/molbev/msq349

Lango Allen, H., Estrada, K., Lettre, G., Berndt, S.I., Weedon, M.N., Rivadeneira, F., Willer, C.J., Jackson, A.U., Vedantam, S., Raychaudhuri, S., Ferreira, T., Wood, A.R., Weyant, R.J., Segrè, A.V., Speliotes, E.K., Wheeler, E., Soranzo, N., Park, J.-H.,

Yang, J., Gudbjartsson, D., Heard-Costa, N.L., Randall, J.C., Qi, L., Vernon Smith, A., Mägi, R., Pastinen, T., Liang, L., Heid, I.M., Luan, J., Thorleifsson, G., Winkler, T.W., Goddard, M.E., Sin Lo, K., Palmer, C., Workalemahu, T., Aulchenko, Y.S., Johansson, A., Zillikens, M.C., Feitosa, M.F., Esko, T., Johnson, T., Ketkar, S., Kraft, P., Mangino, M., Prokopenko, I., Absher, D., Albrecht, E., Ernst, F., Glazer, N.L., Hayward, C., Hottenga, J.-J., Jacobs, K.B., Knowles, J.W., Kutalik, Z., Monda, K.L., Polasek, O., Preuss, M., Rayner, N.W., Robertson, N.R., Steinthorsdottir, V., Tyrer, J.P., Voight, B.F., Wiklund, F., Xu, J., Zhao, J.H., Nyholt, D.R., Pellikka, N., Perola, M., Perry, J.R.B., Surakka, I., Tammesoo, M.-L., Altmaier, E.L., Amin, N., Aspelund, T., Bhangale, T., Boucher, G., Chasman, D.I., Chen, C., Coin, L., Cooper, M.N., Dixon, A.L., Gibson, Q., Grundberg, E., Hao, K., Juhani Junttila, M., Kaplan, L.M., Kettunen, J., König, I.R., Kwan, T., Lawrence, R.W., Levinson, D.F., Lorentzon, M., McKnight, B., Morris, A.P., Müller, M., Suh Ngwa, J., Purcell, S., Rafelt, S., Salem, R.M., Salvi, E., Sanna, S., Shi, J., Sovio, U., Thompson, J.R., Turchin, M.C., Vandenput, L., Verlaan, D.J., Vitart, V., White, C.C., Ziegler, A., Almgren, P., Balmforth, A.J., Campbell, H., Citterio, L., De Grandi, A., Dominiczak, A., Duan, J., Elliott, P., Elosua, R., Eriksson, J.G., Freimer, N.B., Geus, E.J.C., Glorioso, N., Haiqing, S., Hartikainen, A.-L., Havulinna, A.S., Hicks, A.A., Hui, J., Igl, W., Illig, T., Jula, A., Kajantie, E., Kilpeläinen, T.O., Koiranen, M., Kolcic, I., Koskinen, S., Kovacs, P., Laitinen, J., Liu, J., Lokki, M.-L., Marusic, A., Maschio, A., Meitinger, T., Mulas, A., Paré, G., Parker, A.N., Peden, J.F., Petersmann, A., Pichler, I., Pietiläinen, K.H., Pouta, A., Ridderstråle, M., Rotter, J.I., Sambrook, J.G., Sanders, A.R., Schmidt, C.O., Sinisalo, J., Smit, J.H., Stringham, H.M., Bragi Walters, G., Widen, E., Wild, S.H., Willemsen, G., Zagato, L., Zgaga, L., Zitting, P., Alavere, H., Farrall, M., McArdle, W.L., Nelis, M., Peters, M.J., Ripatti, S., van Meurs, J.B.J., Aben, K.K., Ardlie, K.G., Beckmann, J.S., Beilby, J.P., Bergman, R.N., Bergmann, S., Collins, F.S., Cusi, D., Heijer, M. den, Eiriksdottir, G., Gejman, P.V., Hall, A.S., Hamsten, A., Huikuri, H.V., Iribarren, C., Kähönen, M., Kaprio, J., Kathiresan, S., Kiemeney, L., Kocher, T., Launer, L.J., Lehtimäki, T., Melander, O., Mosley, T.H., Musk, A.W., Nieminen, M.S., O'Donnell, C.J., Ohlsson, C., Oostra, B., Palmer, L.J., Raitakari, O., Ridker, P.M., Rioux, J.D., Rissanen, A., Rivolta, C., Schunkert, H., Shuldiner, A.R., Siscovick, D.S., Stumvoll, M., Tönjes, A., Tuomilehto, J., van Ommen, G.-J., Viikari, J., Heath, A.C., Martin, N.G., Montgomery, G.W., Province, M.A., Kayser, M., Arnold, A.M., Atwood, L.D., Boerwinkle, E., Chanock, S.J., Deloukas, P., Gieger, C., Grönberg, H., Hall, P., Hattersley, A.T., Hengstenberg, C., Hoffman, W., Lathrop, G.M., Salomaa, V., Schreiber, S., Uda, M., Waterworth, D., Wright, A.F., Assimes, T.L., Barroso, I., Hofman, A., Mohlke, K.L., Boomsma, D.I., Caulfield, M.J., Cupples, L.A., Erdmann, J., Fox, C.S., Gudnason, V., Gyllensten, U., Harris, T.B., Hayes, R.B., Jarvelin, M.-R., Mooser, V., Munroe, P.B., Ouwehand, W.H., Penninx, B.W., Pramstaller, P.P., Quertermous, T., Rudan, I., Samani, N.J., Spector, T.D., Völzke, H., Watkins, H., Wilson, J.F., Groop, L.C., Haritunians, T., Hu, F.B., Kaplan, R.C., Metspalu, A., North, K.E., Schlessinger, D., Wareham, N.J., Hunter, D.J., O'Connell, J.R., Strachan, D.P., Wichmann, H.-E., Borecki, I.B., van Duijn, C.M., Schadt, E.E., Thorsteinsdottir, U., Peltonen, L., Uitterlinden, A.G., Visscher, P.M., Chatterjee, N., Loos, R.J.F., Boehnke, M., McCarthy, M.I., Ingelsson, E., Lindgren, C.M., Abecasis, G.R., Stefansson, K., Frayling, T.M., Hirschhorn, J.N., 2010. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 467, 832–838. doi:10.1038/nature09410

Lappalainen, T., Sammeth, M., Friedländer, M.R., Hoen, P.A.C. 't, Monlong, J., Rivas, M.A., Gonzàlez-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P.G., Barann, M., Wieland, T., Greger, L., van Iterson, M., Almlöf, J., Ribeca, P., Pulyakhina, I., Esser, D., Giger, T., Tikhonov, A., Sultan, M., Bertier, G., MacArthur, D.G., Lek, M., Lizano, E., Buermans, H.P.J., Padioleau, I., Schwarzmayr, T., Karlberg, O., Ongen, H., Kilpinen, H., Beltran, S., Gut, M., Kahlem, K., Amstislavskiy, V., Stegle, O., Pirinen, M., Montgomery, S.B., Donnelly, P., McCarthy, M.I., Flicek, P., Strom, T.M., Geuvadis Consortium, Lehrach, H., Schreiber, S., Sudbrak, R., Carracedo, A., Antonarakis, S.E., Häsler, R., Syvänen, A.-C., van Ommen, G.-J., Brazma, A., Meitinger, T., Rosenstiel, P., Guigó, R., Gut, I.G., Estivill, X., Dermitzakis, E.T., 2013. Transcriptome and genome sequencing uncovers functional variation in humans. Nature 501, 506–511. doi:10.1038/nature12531

Lau, N.C., Lim, L.P., Weinstein, E.G., Bartel, D.P., 2001. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science 294, 858–862. doi:10.1126/science.1065062

Laval, G., Patin, E., Barreiro, L.B., Quintana-Murci, L., 2010. Formulating a historical and demographic model of recent human evolution based on resequencing data from noncoding regions. PLoS ONE 5, e10284. doi:10.1371/journal.pone.0010284

Lecellier, C.-H., Dunoyer, P., Arar, K., Lehmann-Che, J., Eyquem, S., Himber, C., Saïb, A., Voinnet, O., 2005. A Cellular MicroRNA Mediates Antiviral Defense in Human Cells. Science 308, 557–560. doi:10.1126/science.1108784

Lee, M.N., Ye, C., Villani, A.-C., Raj, T., Li, W., Eisenhaure, T.M., Imboywa, S.H., Chipendo, P.I., Ran, F.A., Slowikowski, K., Ward, L.D., Raddassi, K., McCabe, C., Lee, M.H., Frohlich, I.Y., Hafler, D.A., Kellis, M., Raychaudhuri, S., Zhang, F., Stranger, B.E., Benoist, C.O., Jager, P.L.D., Regev, A., Hacohen, N., 2014. Common Genetic Variants Modulate Pathogen-Sensing Responses in Human Dendritic Cells. Science 343, 1246980. doi:10.1126/science.1246980

Lee, M.S., Kim, Y.-J., 2007. Signaling pathways downstream of pattern-recognition receptors and their cross talk. Annu. Rev. Biochem. 76, 447–480. doi:10.1146/annurev.biochem.76.060605.122847

Lee, R.C., Ambros, V., 2001. An extensive class of small RNAs in Caenorhabditis elegans. Science 294, 862–864. doi:10.1126/science.1065329

Lee, R.C., Feinbaum, R.L., Ambros, V., 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75, 843–854.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Rådmark, O., Kim, S., Kim, V.N., 2003. The nuclear RNase III Drosha initiates microRNA processing. Nature 425, 415–419. doi:10.1038/nature01957

Lee, Y., Kim, M., Han, J., Yeom, K.-H., Lee, S., Baek, S.H., Kim, V.N., 2004. MicroRNA genes are transcribed by RNA polymerase II. EMBO J. 23, 4051–4060. doi:10.1038/sj.emboj.7600385

Lemaitre, B., Hoffmann, J., 2007. The host defense of Drosophila melanogaster. Annu. Rev. Immunol. 25, 697–743. doi:10.1146/annurev.immunol.25.022106.141615

Lemaitre, B., Nicolas, E., Michaut, L., Reichhart, J.M., Hoffmann, J.A., 1996. The dorsoventral regulatory gene cassette spätzle/Toll/cactus controls the potent antifungal response in Drosophila adults. Cell 86, 973–983.

Leslie, T., Briceno, M., Mayan, I., Mohammed, N., Klinkenberg, E., Sibley, C.H., Whitty, C.J., Rowland, M., 2010. The impact of phenotypic and genotypic G6PD deficiency on risk of plasmodium vivax infection: a case-control study amongst Afghan refugees in Pakistan. PLoS medicine 7, e1000283. doi:10.1371/journal.pmed.1000283

Lewontin, R.C., Krakauer, J., 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics 74, 175–195.

Lichten, M., Goldman, A.S.H., 1995. Meiotic Recombination Hotspots. Annual Review of Genetics 29, 423–444. doi:10.1146/annurev.ge.29.120195.002231

Lim, L.P., Lau, N.C., Garrett-Engele, P., Grimson, A., Schelter, J.M., Castle, J., Bartel, D.P., Linsley, P.S., Johnson, J.M., 2005. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature 433, 769–773. doi:10.1038/nature03315

Lin, R., Lacoste, J., Nakhaei, P., Sun, Q., Yang, L., Paz, S., Wilkinson, P., Julkunen, I., Vitour, D., Meurs, E., Hiscott, J., 2006. Dissociation of a MAVS/IPS-1/VISA/Cardif-IKKepsilon molecular complex from the mitochondrial outer membrane by hepatitis C virus NS3-4A proteolytic cleavage. J. Virol. 80, 6072–6083. doi:10.1128/JVI.02495-05

Li, S.-C., Liao, Y.-L., Ho, M.-R., Tsai, K.-W., Lai, C.-H., Lin, W., 2012. miRNA arm selection and isomiR distribution in gastric cancer. BMC Genomics 13 Suppl 1, S13. doi:10.1186/1471-2164-13-S1-S13

Liston, A., Papadopoulou, A.S., Danso-Abeam, D., Dooley, J., 2012. MicroRNA-29 in the adaptive immune system: setting the threshold. Cell. Mol. Life Sci. 69, 3533–3541. doi:10.1007/s00018-012-1124-0

Litman, G.W., Cannon, J.P., Dishaw, L.J., 2005. Reconstructing immune phylogeny: new perspectives. Nat. Rev. Immunol. 5, 866–879. doi:10.1038/nri1712

Liu, Y., Wang, X., Jiang, J., Cao, Z., Yang, B., Cheng, X., 2011. Modulation of T cell cytokine production by miR-144* with elevated expression in patients with pulmonary tuberculosis. Mol. Immunol. 48, 1084–1090. doi:10.1016/j.molimm.2011.02.001

Li, W.H., Sadler, L.A., 1991. Low nucleotide diversity in man. Genetics 129, 513–523.

Li, X.-D., Wu, J., Gao, D., Wang, H., Sun, L., Chen, Z.J., 2013. Pivotal roles of cGAS-cGAMP signaling in antiviral defense and immune adjuvant effects. Science 341, 1390–1394. doi:10.1126/science.1244040

Li, Y., Shi, X., 2013. MicroRNAs in the regulation of TLR and RIG-I pathways. Cell Mol Immunol 10, 65–71. doi:10.1038/cmi.2012.55

Lodish, H.F., Zhou, B., Liu, G., Chen, C.-Z., 2008. Micromanagement of the immune system by microRNAs. Nat. Rev. Immunol. 8, 120–130. doi:10.1038/nri2252

Lohmueller, K.E., Albrechtsen, A., Li, Y., Kim, S.Y., Korneliussen, T., Vinckenbosch, N., Tian, G., Huerta-Sanchez, E., Feder, A.F., Grarup, N., Jørgensen, T., Jiang, T., Witte, D.R., Sandbæk, A., Hellmann, I., Lauritzen, T., Hansen, T., Pedersen, O., Wang, J., Nielsen, R., 2011. Natural selection affects multiple aspects of genetic variation at putatively neutral sites across the human genome. PLoS Genet. 7, e1002326. doi:10.1371/journal.pgen.1002326

Lönnroth, K., Jaramillo, E., Williams, B.G., Dye, C., Raviglione, M., 2009. Drivers of tuberculosis epidemics: The role of risk factors and social determinants. Social Science & Medicine 68, 2240–2246. doi:10.1016/j.socscimed.2009.03.041

Loo, Y.-M., Gale, M., 2011. Immune signaling by RIG-I-like receptors. Immunity 34, 680–692. doi:10.1016/j.immuni.2011.05.003

Louicharoen, C., Patin, E., Paul, R., Nuchprayoon, I., Witoonpanich, B., Peerapittayamongkol, C., Casademont, I., Sura, T., Laird, N.M., Singhasivanon, P., Quintana-Murci, L., Sakuntabhai, A., 2009. Positively selected G6PD-Mahidol mutation reduces Plasmodium vivax density in Southeast Asians. Science 326, 1546–1549. doi:10.1126/science.1178849

Lu, H., Lu, N., Weng, L., Yuan, B., Liu, Y.-J., Zhang, Z., 2014. DHX15 senses double-stranded RNA in myeloid dendritic cells. J. Immunol. 193, 1364–1372. doi:10.4049/jimmunol.1303322

Luisi, P., Alvarez-Ponce, D., Dall'Olio, G.M., Sikora, M., Bertranpetit, J., Laayouni, H., 2012. Network-Level and Population Genetics Analysis of the Insulin/TOR Signal Transduction Pathway Across Human Populations. Mol Biol Evol 29, 1379–1392. doi:10.1093/molbev/msr298

Luisi, P., Alvarez-Ponce, D., Pybus, M., Fares, M.A., Bertranpetit, J., Laayouni, H., 2015. Recent Positive Selection Has Acted on Genes Encoding Proteins with More Interactions within the Whole Human Interactome. Genome Biol Evol 7, 1141–1154. doi:10.1093/gbe/evv055

Luo, D., Kohlway, A., Pyle, A.M., 2013. Duplex RNA activated ATPases (DRAs): platforms for RNA sensing, signaling and processing. RNA Biol 10, 111–120. doi:10.4161/rna.22706

Lytle, J.R., Yario, T.A., Steitz, J.A., 2007. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR. Proc. Natl. Acad. Sci. U.S.A. 104, 9667–9672. doi:10.1073/pnas.0703820104

Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Meehan, W., Blackburn, J., Semino, O., Scozzari, R., Cruciani, F., Taha, A., Shaari, N.K., Raja, J.M., Ismail, P., Zainuddin, Z., Goodwin, W., Bulbeck, D., Bandelt, H.-J., Oppenheimer, S., Torroni, A., Richards, M., 2005. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. Science 308, 1034–1036. doi:10.1126/science.1109792

Maertzdorf, J., Weiner, J., Mollenkopf, H.-J., TBornotTB Network, Bauer, T., Prasse, A., Müller-Quernheim, J., Kaufmann, S.H.E., 2012. Common patterns and disease-related signatures in tuberculosis and sarcoidosis. Proc. Natl. Acad. Sci. U.S.A. 109, 7853–7858. doi:10.1073/pnas.1121072109

Ma, F., Xu, S., Liu, X., Zhang, Q., Xu, X., Liu, M., Hua, M., Li, N., Yao, H., Cao, X., 2011. The microRNA miR-29 controls innate and adaptive immune responses to intracellular bacterial infection by targeting interferon-γ. Nat. Immunol. 12, 861–869. doi:10.1038/ni.2073

Majewski, J., Pastinen, T., 2011. The study of eQTL variations by RNA-seq: from SNPs to phenotypes. Trends Genet. 27, 72–79. doi:10.1016/j.tig.2010.10.006

Manry, J., Laval, G., Patin, E., Fornarino, S., Itan, Y., Fumagalli, M., Sironi, M., Tichit, M., Bouchier, C., Casanova, J.-L., Barreiro, L.B., Quintana-Murci, L., 2011a. Evolutionary genetic dissection of human interferons. J. Exp. Med. 208, 2747–2759. doi:10.1084/jem.20111680

Manry, J., Laval, G., Patin, E., Fornarino, S., Tichit, M., Bouchier, C., Barreiro, L.B., Quintana-Murci, L., 2011b. Evolutionary genetics evidence of an essential, nonredundant role of the IFN-γ pathway in protective immunity. Hum. Mutat. 32, 633–642. doi:10.1002/humu.21484

Maranville, J.C., Luca, F., Richards, A.L., Wen, X., Witonsky, D.B., Baxter, S., Stephens, M., Di Rienzo, A., 2011. Interactions between glucocorticoid treatment and cis-regulatory polymorphisms contribute to cellular response phenotypes. PLoS Genet. 7, e1002162. doi:10.1371/journal.pgen.1002162

Maranville, J.C., Luca, F., Stephens, M., Di Rienzo, A., 2012. Mapping gene-environment interactions at regulatory polymorphisms: insights into mechanisms of phenotypic variation. Transcription 3, 56–62. doi:10.4161/trns.19497

Mariathasan, S., Weiss, D.S., Newton, K., McBride, J., O'Rourke, K., Roose-Girma, M., Lee, W.P., Weinrauch, Y., Monack, D.M., Dixit, V.M., 2006. Cryopyrin activates the inflammasome in response to toxins and ATP. Nature 440, 228–232. doi:10.1038/nature04515

Marshak-Rothstein, A., 2006. Toll-like receptors in systemic autoimmune disease. Nat. Rev. Immunol. 6, 823–835. doi:10.1038/nri1957

Martinez, N.J., Ow, M.C., Reece-Hoyes, J.S., Barrasa, M.I., Ambros, V.R., Walhout, A.J.M., 2008. Genome-scale spatiotemporal analysis of Caenorhabditis elegans microRNA promoter activity. Genome Res. 18, 2005–2015. doi:10.1101/gr.083055.108

Martinon, F., Burns, K., Tschopp, J., 2002. The inflammasome: a molecular platform triggering activation of inflammatory caspases and processing of proIL-beta. Mol. Cell 10, 417–426.

Ma, Y., Haynes, R.L., Sidman, R.L., Vartanian, T., 2007. TLR8: an innate immune receptor in brain, neurons and axons. Cell Cycle 6, 2859–2868.

Ma, Y., Li, J., Chiu, I., Wang, Y., Sloane, J.A., Lü, J., Kosaras, B., Sidman, R.L., Volpe, J.J., Vartanian, T., 2006. Toll-like receptor 8 functions as a negative regulator of neurite outgrowth and inducer of neuronal apoptosis. J. Cell Biol. 175, 209–215. doi:10.1083/jcb.200606016

McDonald, J.H., Kreitman, M., 1991. Adaptive protein evolution at the Adh locus in Drosophila. Nature 351, 652–654. doi:10.1038/351652a0

McLaren, P.J., Coulonges, C., Ripke, S., van den Berg, L., Buchbinder, S., Carrington, M., Cossarizza, A., Dalmau, J., Deeks, S.G., Delaneau, O., De Luca, A., Goedert, J.J., Haas, D., Herbeck, J.T., Kathiresan, S., Kirk, G.D., Lambotte, O., Luo, M., Mallal, S., van Manen, D., Martinez-Picado, J., Meyer, L., Miro, J.M., Mullins, J.I., Obel, N., O'Brien, S.J., Pereyra, F., Plummer, F.A., Poli, G., Qi, Y., Rucart, P., Sandhu, M.S., Shea, P.R., Schuitemaker, H., Theodorou, I., Vannberg, F., Veldink, J., Walker, B.D., Weintrob, A., Winkler, C.A., Wolinsky, S., Telenti, A., Goldstein, D.B., de Bakker, P.I.W., Zagury, J.-F., Fellay, J., 2013. Association study of common genetic variants and HIV-1 acquisition in 6,300 infected cases and 7,200 controls. PLoS Pathog. 9, e1003515. doi:10.1371/journal.ppat.1003515

McVean, G.A., Myers, S.R., Hunt, S., Deloukas, P., Bentley, D.R., Donnelly, P., 2004. The fine-scale structure of recombination rate variation in the human genome. Science 304, 581–4. doi:10.1126/science.1092500

Medzhitov, R., 2001. Toll-like receptors and innate immunity. Nat. Rev. Immunol. 1, 135–145. doi:10.1038/35100529

Medzhitov, R., Janeway, C.A., 1998. Innate immune recognition and control of adaptive immune responses. Seminars in Immunology 10, 351–353. doi:10.1006/smim.1998.0136

Medzhitov, R., Preston-Hurlburt, P., Janeway, C.A., 1997. A human homologue of the Drosophila Toll protein signals activation of adaptive immunity. Nature 388, 394–397. doi:10.1038/41131

Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., Tuschl, T., 2004. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. Mol. Cell 15, 185–197. doi:10.1016/j.molcel.2004.07.007

Mendez, F.L., Watkins, J.C., Hammer, M.F., 2013. Neandertal Origin of Genetic Variation at the Cluster of OAS Immunity Genes. Mol Biol Evol 30, 798–801. doi:10.1093/molbev/mst004

Mendez, F.L., Watkins, J.C., Hammer, M.F., 2012a. Global Genetic Variation at OAS1 Provides Evidence of Archaic Admixture in Melanesian Populations. Mol Biol Evol 29, 1513–1520. doi:10.1093/molbev/msr301

Mendez, F.L., Watkins, J.C., Hammer, M.F., 2012b. A Haplotype at STAT2 Introgressed from Neanderthals and Serves as a Candidate of Positive Selection in Papua New Guinea. The American Journal of Human Genetics 91, 265–274. doi:10.1016/j.ajhg.2012.06.015

Mendizabal, I., Marigorta, U.M., Lao, O., Comas, D., 2012. Adaptive evolution of loci covarying with the human African Pygmy phenotype. Hum. Genet. 131, 1305–1317. doi:10.1007/s00439-012-1157-3

Messer, P.W., Petrov, D.A., 2013. Population genomics of rapid adaptation by soft selective sweeps. Trends in Ecology & Evolution 28, 659–669. doi:10.1016/j.tree.2013.08.003

Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., Mallick, S., Schraiber, J.G., Jay, F., Prüfer, K., de Filippo, C., Sudmant, P.H., Alkan, C., Fu, Q., Do, R., Rohland, N., Tandon, A., Siebauer, M., Green, R.E., Bryc, K., Briggs, A.W., Stenzel, U., Dabney, J., Shendure, J., Kitzman, J., Hammer, M.F., Shunkov, M.V., Derevianko, A.P., Patterson, N., Andrés, A.M., Eichler, E.E., Slatkin, M., Reich, D., Kelso, J., Pääbo, S., 2012. A high-coverage genome sequence from an archaic Denisovan individual. Science 338, 222–226. doi:10.1126/science.1224344

Meyers, B.C., Shen, K.A., Rohani, P., Gaut, B.S., Michelmore, R.W., 1998. Receptor-like genes in the major resistance locus of lettuce are subject to divergent selection. Plant Cell 10, 1833–1846.

Miller, L.H., Mason, S.J., Clyde, D.F., McGinniss, M.H., 1976. The resistance factor to Plasmodium vivax in blacks. The Duffy-blood-group genotype, FyFy. The New England journal of medicine 295, 302–4. doi:10.1056/NEJM197608052950602

Mira, M.T., Alcaïs, A., Nguyen, V.T., Moraes, M.O., Di Flumeri, C., Vu, H.T., Mai, C.P., Nguyen, T.H., Nguyen, N.B., Pham, X.K., Sarno, E.N., Alter, A., Montpetit, A.,

Moraes, M.E., Moraes, J.R., Doré, C., Gallant, C.J., Lepage, P., Verner, A., Van De Vosse, E., Hudson, T.J., Abel, L., Schurr, E., 2004. Susceptibility to leprosy is associated with PARK2 and PACRG. Nature 427, 636–640. doi:10.1038/nature02326

Misch, E.A., Macdonald, M., Ranjit, C., Sapkota, B.R., Wells, R.D., Siddiqui, M.R., Kaplan, G., Hawn, T.R., 2008. Human TLR1 deficiency is associated with impaired mycobacterial signaling and protection from leprosy reversal reaction. PLoS Negl Trop Dis 2, e231. doi:10.1371/journal.pntd.0000231

Möller, M., Hoal, E.G., 2010. Current findings, challenges and novel approaches in human genetic susceptibility to tuberculosis. Tuberculosis 90, 71–83. doi:10.1016/j.tube.2010.02.002

Mondragón-Palomino, M., Meyers, B.C., Michelmore, R.W., Gaut, B.S., 2002. Patterns of positive selection in the complete NBS-LRR gene family of Arabidopsis thaliana. Genome Res. 12, 1305–1315. doi:10.1101/gr.159402

Montagner, S., Orlandi, E.M., Merante, S., Monticelli, S., 2013. The role of miRNAs in mast cells and other innate immune cells. Immunol. Rev. 253, 12–24. doi:10.1111/imr.12042

Montecalvo, A., Larregina, A.T., Shufesky, W.J., Stolz, D.B., Sullivan, M.L.G., Karlsson, J.M., Baty, C.J., Gibson, G.A., Erdos, G., Wang, Z., Milosevic, J., Tkacheva, O.A., Divito, S.J., Jordan, R., Lyons-Weiler, J., Watkins, S.C., Morelli, A.E., 2012. Mechanism of transfer of functional microRNAs between mouse dendritic cells via exosomes. Blood 119, 756–766. doi:10.1182/blood-2011-02-338004

Montgomery, S.B., Dermitzakis, E.T., 2011. From expression QTLs to personalized transcriptomics. Nat. Rev. Genet. 12, 277–282. doi:10.1038/nrg2969

Moore, C.B., Bergstralh, D.T., Duncan, J.A., Lei, Y., Morrison, T.E., Zimmermann, A.G., Accavitti-Loper, M.A., Madden, V.J., Sun, L., Ye, Z., Lich, J.D., Heise, M.T., Chen, Z., Ting, J.P.-Y., 2008. NLRX1 is a regulator of mitochondrial antiviral immunity. Nature 451, 573–577. doi:10.1038/nature06501

Mukherjee, S., Sarkar-Roy, N., Wagener, D.K., Majumder, P.P., 2009. Signatures of natural selection are not uniform across genes of innate immune system, but purifying selection is the dominant signature. Proc. Natl. Acad. Sci. U.S.A. 106, 7073–7078. doi:10.1073/pnas.0811357106

Mullighan, C.G., Heatley, S., Doherty, K., Szabo, F., Grigg, A., Hughes, T.P., Schwarer, A.P., Szer, J., Tait, B.D., Bik To, L., Bardy, P.G., 2002. Mannose-binding lectin gene polymorphisms are associated with major infection following allogeneic hemopoietic stem cell transplantation. Blood 99, 3524–3529.

Myers, S., Freeman, C., Auton, A., Donnelly, P., McVean, G., 2008. A common sequence motif associated with recombination hot spots and genome instability in humans. Nat. Genet. 40, 1124–1129. doi:10.1038/ng.213

Nakajima, T., Ohtani, H., Satta, Y., Uno, Y., Akari, H., Ishida, T., Kimura, A., 2008. Natural selection in the TLR-related genes in the course of primate evolution. Immunogenetics 60, 727–735. doi:10.1007/s00251-008-0332-0

Nakaoka, H., Inoue, I., 2015. Distribution of HLA haplotypes across Japanese Archipelago: similarity, difference and admixture. J. Hum. Genet. doi:10.1038/jhg.2015.90

Naranbhai, V., Fairfax, B.P., Makino, S., Humburg, P., Wong, D., Ng, E., Hill, A.V.S., Knight, J.C., 2015. Genomic modulators of gene expression in human neutrophils. Nat Commun 6, 7545. doi:10.1038/ncomms8545

Nei, M., 1987. Molecular evolutionary genetics. Columbia University Press, New York.

Newport, M.J., Huxley, C.M., Huston, S., Hawrylowicz, C.M., Oostra, B.A., Williamson, R., Levin, M., 1996. A mutation in the interferon-gamma-receptor gene and susceptibility to mycobacterial infection. N. Engl. J. Med. 335, 1941–1949. doi:10.1056/NEJM199612263352602

Nica, A.C., Montgomery, S.B., Dimas, A.S., Stranger, B.E., Beazley, C., Barroso, I., Dermitzakis, E.T., 2010. Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. PLoS Genet. 6, e1000895. doi:10.1371/journal.pgen.1000895

Nicolae, D.L., Gamazon, E., Zhang, W., Duan, S., Dolan, M.E., Cox, N.J., 2010. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet. 6, e1000888. doi:10.1371/journal.pgen.1000888

Nielsen, R., 2005. Molecular signatures of natural selection. Annu. Rev. Genet. 39, 197–218. doi:10.1146/annurev.genet.39.073003.112420

Nielsen, R., Bustamante, C., Clark, A.G., Glanowski, S., Sackton, T.B., Hubisz, M.J., Fledel-Alon, A., Tanenbaum, D.M., Civello, D., White, T.J., J Sninsky, J., Adams, M.D., Cargill, M., 2005a. A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol. 3, e170. doi:10.1371/journal.pbio.0030170

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M.J., Clark, A.G., Bustamante, C., 2005b. Genomic scans for selective sweeps using SNP data. Genome Res. 15, 1566–1575. doi:10.1101/gr.4252305

Niu, M., Tabari, E.S., Su, Z., 2014. De novo prediction of cis-regulatory elements and modules through integrative analysis of a large number of ChIP datasets. BMC Genomics 15, 1047. doi:10.1186/1471-2164-15-1047

Norton, H.L., Kittles, R.A., Parra, E., McKeigue, P., Mao, X., Cheng, K., Canfield, V.A., Bradley, D.G., McEvoy, B., Shriver, M.D., 2007. Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. Molecular biology and evolution 24, 710–22. doi:10.1093/molbev/msl203

O'Connell, R.M., Rao, D.S., Baltimore, D., 2012. microRNA regulation of inflammatory responses. Annu. Rev. Immunol. 30, 295–312. doi:10.1146/annurev-immunol-020711-075013

O'Connell, R.M., Rao, D.S., Chaudhuri, A.A., Baltimore, D., 2010. Physiological and pathological roles for microRNAs in the immune system. Nat. Rev. Immunol. 10, 111–122. doi:10.1038/nri2708

Ogura, Y., Inohara, N., Benito, A., Chen, F.F., Yamaoka, S., Nunez, G., 2001. Nod2, a Nod1/Apaf-1 family member that is restricted to monocytes and activates NF-kappaB. J. Biol. Chem. 276, 4812–4818. doi:10.1074/jbc.M008072200

Ohman, T., Rintahaka, J., Kalkkinen, N., Matikainen, S., Nyman, T.A., 2009. Actin and RIG-I/MAVS signaling components translocate to mitochondria upon influenza A virus infection of human primary macrophages. J. Immunol. 182, 5682–5692. doi:10.4049/jimmunol.0803093

Okada, S., Markle, J.G., Deenick, E.K., Mele, F., Averbuch, D., Lagos, M., Alzahrani, M., Al-Muhsen, S., Halwani, R., Ma, C.S., Wong, N., Soudais, C., Henderson, L.A., Marzouqa, H., Shamma, J., Gonzalez, M., Martinez-Barricarte, R., Okada, C., Avery, D.T., Latorre, D., Deswarte, C., Jabot-Hanin, F., Torrado, E., Fountain, J., Belkadi, A., Itan, Y., Boisson, B., Migaud, M., Arlehamn, C.S.L., Sette, A., Breton, S., McCluskey, J., Rossjohn, J., de Villartay, J.-P., Moshous, D., Hambleton, S., Latour, S., Arkwright, P.D., Picard, C., Lantz, O., Engelhard, D., Kobayashi, M., Abel, L., Cooper, A.M., Notarangelo, L.D., Boisson-Dupuis, S., Puel, A., Sallusto, F., Bustamante, J., Tangye, S.G., Casanova, J.-L., 2015. IMMUNODEFICIENCIES. Impairment of immunity to Candida and Mycobacterium in humans with bi-allelic RORC mutations. Science 349, 606–613. doi:10.1126/science.aaa4282

Okamura, K., Hagen, J.W., Duan, H., Tyler, D.M., Lai, E.C., 2007. The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. Cell 130, 89–100. doi:10.1016/j.cell.2007.06.028

Olsen, P.H., Ambros, V., 1999. The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. Dev. Biol. 216, 671–680. doi:10.1006/dbio.1999.9523

O'Neill, L.A., Sheedy, F.J., McCoy, C.E., 2011. MicroRNAs: the fine-tuners of Toll-like receptor signalling. Nat Rev Immunol 11, 163–175. doi:10.1038/nri2957

Onoguchi, K., Yoneyama, M., Takemura, A., Akira, S., Taniguchi, T., Namiki, H., Fujita, T., 2007. Viral infections activate types I and III interferon genes through a common mechanism. J. Biol. Chem. 282, 7576–7581. doi:10.1074/jbc.M608618200

Oosting, M., Cheng, S.-C., Bolscher, J.M., Vestering-Stenger, R., Plantinga, T.S., Verschueren, I.C., Arts, P., Garritsen, A., van Eenennaam, H., Sturm, P., Kullberg, B.-J., Hoischen, A., Adema, G.J., van der Meer, J.W.M., Netea, M.G., Joosten, L.A.B., 2014. Human TLR10 is an anti-inflammatory pattern-recognition receptor. Proc. Natl. Acad. Sci. U.S.A. 111, E4478–4484. doi:10.1073/pnas.1410293111

Ortiz, M., Kaessmann, H., Zhang, K., Bashirova, A., Carrington, M., Quintana-Murci, L., Telenti, A., 2008. The evolutionary history of the CD209 (DC-SIGN) family in humans and non-human primates. Genes Immun. 9, 483–492. doi:10.1038/gene.2008.40

Osterlund, P., Veckman, V., Sirén, J., Klucher, K.M., Hiscott, J., Matikainen, S., Julkunen, I., 2005. Gene expression and antiviral activity of alpha/beta interferons and interleukin-29 in virus-infected human myeloid dendritic cells. J. Virol. 79, 9608–9617. doi:10.1128/JVI.79.15.9608-9617.2005

Pannicke, U., Baumann, B., Fuchs, S., Henneke, P., Rensing-Ehl, A., Rizzi, M., Janda, A., Hese, K., Schlesier, M., Holzmann, K., Borte, S., Laux, C., Rump, E.-M., Rosenberg, A., Zelinski, T., Schrezenmeier, H., Wirth, T., Ehl, S., Schroeder, M.L., Schwarz, K., 2013. Deficiency of innate and acquired immunity caused by an IKBKB mutation. N. Engl. J. Med. 369, 2504–2514. doi:10.1056/NEJMoa1309199

Park, J.-E., Heo, I., Tian, Y., Simanshu, D.K., Chang, H., Jee, D., Patel, D.J., Kim, V.N., 2011. Dicer recognizes the 5' end of RNA for efficient and accurate processing. Nature 475, 201–205. doi:10.1038/nature10198

Parniske, M., Hammond-Kosack, K.E., Golstein, C., Thomas, C.M., Jones, D.A., Harrison, K., Wulff, B.B., Jones, J.D., 1997. Novel disease resistance specificities result from

sequence exchange between tandemly repeated genes at the Cf-4/9 locus of tomato. Cell 91, 821–832.

Parts, L., Hedman, Å.K., Keildson, S., Knights, A.J., Abreu-Goodger, C., van de Bunt, M., Guerra-Assunção, J.A., Bartonicek, N., van Dongen, S., Mägi, R., Nisbet, J., Barrett, A., Rantalainen, M., Nica, A.C., Quail, M.A., Small, K.S., Glass, D., Enright, A.J., Winn, J., MuTHER Consortium, Deloukas, P., Dermitzakis, E.T., McCarthy, M.I., Spector, T.D., Durbin, R., Lindgren, C.M., 2012. Extent, causes, and consequences of small RNA expression variation in human adipose tissue. PLoS Genet. 8, e1002704. doi:10.1371/journal.pgen.1002704

Parvatiyar, K., Zhang, Z., Teles, R.M., Ouyang, S., Jiang, Y., Iyer, S.S., Zaver, S.A., Schenk, M., Zeng, S., Zhong, W., Liu, Z.-J., Modlin, R.L., Liu, Y., Cheng, G., 2012. The helicase DDX41 recognizes the bacterial secondary messengers cyclic di-GMP and cyclic di-AMP to activate a type I interferon immune response. Nat. Immunol. 13, 1155–1161. doi:10.1038/ni.2460

Pasteur, L., Gauthier-Villars, 1870. Etudes sur la maladie des vers a soie, moyen pratique assure de la combattre et d'en prevenir le retour: La pébrine et la flacherie. Gauthier-Villars.

Pedersen, I.M., Cheng, G., Wieland, S., Volinia, S., Croce, C.M., Chisari, F.V., David, M., 2007. Interferon modulation of cellular microRNAs as an antiviral mechanism. Nature 449, 919–922. doi:10.1038/nature06205

Peltonen, L., Pekkarinen, P., Aaltonen, J., 1995. Messages from an isolate: lessons from the Finnish gene pool. Biol. Chem. Hoppe-Seyler 376, 697–704.

Pennings, P.S., Hermisson, J., 2006a. Soft sweeps II--molecular population genetics of adaptation from recurrent mutation or migration. Mol. Biol. Evol. 23, 1076–1084. doi:10.1093/molbev/msj117

Pennings, P.S., Hermisson, J., 2006b. Soft sweeps III: the signature of positive selection from recurrent mutation. PLoS Genet. 2, e186. doi:10.1371/journal.pgen.0020186

Pérez de Diego, R., Sancho-Shimizu, V., Lorenzo, L., Puel, A., Plancoulaine, S., Picard, C., Herman, M., Cardon, A., Durandy, A., Bustamante, J., Vallabhapurapu, S., Bravo, J., Warnatz, K., Chaix, Y., Cascarrigny, F., Lebon, P., Rozenberg, F., Karin, M., Tardieu, M., Al-Muhsen, S., Jouanguy, E., Zhang, S.-Y., Abel, L., Casanova, J.-L., 2010. Human TRAF3 adaptor molecule deficiency leads to impaired Toll-like receptor 3 response and susceptibility to herpes simplex encephalitis. Immunity 33, 400–411. doi:10.1016/j.immuni.2010.08.014

Peters, L., Meister, G., 2007. Argonaute proteins: mediators of RNA silencing. Mol. Cell 26, 611–623. doi:10.1016/j.molcel.2007.05.001

Peterslund, N.A., Koch, C., Jensenius, J.C., Thiel, S., 2001. Association between deficiency of mannose-binding lectin and severe infections after chemotherapy. Lancet 358, 637–638. doi:10.1016/S0140-6736(01)05785-3

Pickrell, J.K., Coop, G., Novembre, J., Kudaravalli, S., Li, J.Z., Absher, D., Srinivasan, B.S., Barsh, G.S., Myers, R.M., Feldman, M.W., Pritchard, J.K., 2009. Signals of recent positive selection in a worldwide sample of human populations. Genome research 19, 826–37. doi:10.1101/gr.087577.108

Pickrell, J.K., Marioni, J.C., Pai, A.A., Degner, J.F., Engelhardt, B.E., Nkadori, E., Veyrieras, J.-B., Stephens, M., Gilad, Y., Pritchard, J.K., 2010. Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature 464, 768–772. doi:10.1038/nature08872

Pisitkun, P., Deane, J.A., Difilippantonio, M.J., Tarasenko, T., Satterthwaite, A.B., Bolland, S., 2006. Autoreactive B cell responses to RNA-related antigens due to TLR7 gene duplication. Science 312, 1669–1672. doi:10.1126/science.1124978

Plagnol, V., Wall, J.D., 2006. Possible ancestral structure in human populations. PLoS Genet. 2, e105. doi:10.1371/journal.pgen.0020105

Poeck, H., Bscheider, M., Gross, O., Finger, K., Roth, S., Rebsamen, M., Hannesschläger, N., Schlee, M., Rothenfusser, S., Barchet, W., Kato, H., Akira, S., Inoue, S., Endres, S., Peschel, C., Hartmann, G., Hornung, V., Ruland, J., 2010. Recognition of RNA virus by RIG-I results in activation of CARD9 and inflammasome signaling for interleukin 1 beta production. Nat. Immunol. 11, 63–69. doi:10.1038/ni.1824

Pritchard, J.K., Cox, N.J., 2002. The allelic architecture of human disease genes: common disease-common variant...or not? Human molecular genetics 11, 2417–23.

Pritchard, J.K., Di Rienzo, A., 2010. Adaptation – not by sweeps alone. Nat Rev Genet 11, 665–667. doi:10.1038/nrg2880

Pritchard, J.K., Pickrell, J.K., Coop, G., 2010. The Genetics of Human Adaptation: Hard Sweeps, Soft Sweeps, and Polygenic Adaptation. Current Biology 20, R208–R215. doi:10.1016/j.cub.2009.11.055

Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P.H., de Filippo, C., Li, H., Mallick, S., Dannemann, M., Fu, Q., Kircher, M., Kuhlwilm, M., Lachmann, M., Meyer, M., Ongyerth, M., Siebauer, M., Theunert, C., Tandon, A., Moorjani, P., Pickrell, J., Mullikin, J.C., Vohr, S.H., Green, R.E., Hellmann, I., Johnson, P.L.F., Blanche, H., Cann, H., Kitzman, J.O., Shendure, J., Eichler, E.E., Lein, E.S., Bakken, T.E., Golovanova, L.V., Doronichev, V.B., Shunkov, M.V., Derevianko, A.P., Viola, B., Slatkin, M., Reich, D., Kelso, J., Pääbo, S., 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature 505, 43–49. doi:10.1038/nature12886

Prugnolle, F., Manica, A., Charpentier, M., Guégan, J.F., Guernier, V., Balloux, F., 2005. Pathogen-Driven Selection and Worldwide HLA Class I Diversity. Current Biology 15, 1022–1027. doi:10.1016/j.cub.2005.04.050

Przeworski, M., Coop, G., Wall, J.D., 2005. The signature of positive selection on standing genetic variation. Evolution 59, 2312–2323.

Qin, P., Stoneking, M., 2015. Denisovan Ancestry in East Eurasian and Native American Populations. Mol. Biol. Evol. doi:10.1093/molbev/msv141

Qrafli, M., Amar, Y., Bourkadi, J., Ben Amor, J., Iraki, G., Bakri, Y., Amzazi, S., Lahlou, O., Seghrouchni, F., Aouad, R. El, Sadki, K., 2014. The CYP7A1 gene rs3808607 variant is associated with susceptibility of tuberculosis in Moroccan population. Pan Afr Med J 18, 1. doi:10.11604/pamj.2014.18.1.3397

Quach, H., Barreiro, L.B., Laval, G., Zidane, N., Patin, E., Kidd, K.K., Kidd, J.R., Bouchier, C., Veuille, M., Antoniewski, C., Quintana-Murci, L., 2009. Signatures of purifying

and local positive selection in human miRNAs. Am. J. Hum. Genet. 84, 316–327. doi:10.1016/j.ajhg.2009.01.022

Quintana-Murci, L., Clark, A.G., 2013. Population genetic tools for dissecting innate immunity in humans. Nat. Rev. Immunol. 13, 280–293. doi:10.1038/nri3421

Quintana-Murci, L., Semino, O., Bandelt, H.J., Passarino, G., McElreavey, K., Santachiara-Benerecetti, A.S., 1999. Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa. Nat. Genet. 23, 437–441. doi:10.1038/70550

Racimo, F., Sankararaman, S., Nielsen, R., Huerta-Sánchez, E., 2015. Evidence for archaic adaptive introgression in humans. Nat. Rev. Genet. 16, 359–371. doi:10.1038/nrg3936

Rajaram, M.V.S., Ni, B., Morris, J.D., Brooks, M.N., Carlson, T.K., Bakthavachalu, B., Schoenberg, D.R., Torrelles, J.B., Schlesinger, L.S., 2011. Mycobacterium tuberculosis lipomannan blocks TNF biosynthesis by regulating macrophage MAPK-activated protein kinase 2 (MK2) and microRNA miR-125b. Proc. Natl. Acad. Sci. U.S.A. 108, 17408–17413. doi:10.1073/pnas.1112660108

Raj, T., Kuchroo, M., Replogle, J.M., Raychaudhuri, S., Stranger, B.E., De Jager, P.L., 2013. Common risk alleles for inflammatory diseases are targets of recent positive selection. Am. J. Hum. Genet. 92, 517–529. doi:10.1016/j.ajhg.2013.03.001

Rast, J.P., Smith, L.C., Loza-Coll, M., Hibino, T., Litman, G.W., 2006. Genomic insights into the immune system of the sea urchin. Science 314, 952–956. doi:10.1126/science.1134301

Reich, D., Green, R.E., Kircher, M., Krause, J., Patterson, N., Durand, E.Y., Viola, B., Briggs, A.W., Stenzel, U., Johnson, P.L.F., Maricic, T., Good, J.M., Marques-Bonet, T., Alkan, C., Fu, Q., Mallick, S., Li, H., Meyer, M., Eichler, E.E., Stoneking, M., Richards, M., Talamo, S., Shunkov, M.V., Derevianko, A.P., Hublin, J.-J., Kelso, J., Slatkin, M., Pääbo, S., 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature 468, 1053–1060. doi:10.1038/nature09710

Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., Ruvkun, G., 2000. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature 403, 901–906. doi:10.1038/35002607

Ricklin, D., Hajishengallis, G., Yang, K., Lambris, J.D., 2010. Complement: a key system for immune surveillance and homeostasis. Nat. Immunol. 11, 785–797. doi:10.1038/ni.1923

Roberts, C.A., Pfister, L.-A., Mays, S., 2009. Letter to the editor: Was tuberculosis present in *Homo erectus* in Turkey? American Journal of Physical Anthropology 139, 442–444. doi:10.1002/ajpa.21056

Rock, F.L., Hardiman, G., Timans, J.C., Kastelein, R.A., Bazan, J.F., 1998. A family of human receptors structurally related to Drosophila Toll. Proc. Natl. Acad. Sci. U.S.A. 95, 588–593.

Rodriguez, A., Griffiths-Jones, S., Ashurst, J.L., Bradley, A., 2004. Identification of mammalian microRNA host genes and transcription units. Genome Res. 14, 1902–1910. doi:10.1101/gr.2722704

Romanoski, C.E., Lee, S., Kim, M.J., Ingram-Drake, L., Plaisier, C.L., Yordanova, R., Tilford, C., Guan, B., He, A., Gargalovic, P.S., Kirchgessner, T.G., Berliner, J.A., Lusis, A.J., 2010. Systems genetics analysis of gene-by-environment interactions in human cells. Am. J. Hum. Genet. 86, 399–410. doi:10.1016/j.ajhg.2010.02.002

Roy, S., Knox, K., Segal, S., Griffiths, D., Moore, C.E., Welsh, K.I., Smarason, A., Day, N.P., McPheat, W.L., Crook, D.W., Hill, A.V.S., Oxford Pneumoccocal Surveillance Group, 2002. MBL genotype and risk of invasive pneumococcal disease: a case-control study. Lancet 359, 1569–1573. doi:10.1016/S0140-6736(02)08516-1

Ruby, J.G., Jan, C.H., Bartel, D.P., 2007. Intronic microRNA precursors that bypass Drosha processing. Nature 448, 83–86. doi:10.1038/nature05983

Ruwende, C., Khoo, S.C., Snow, R.W., Yates, S.N., Kwiatkowski, D., Gupta, S., Warn, P., Allsopp, C.E., Gilbert, S.C., Peschu, N., et al, 1995. Natural selection of hemi- and heterozygotes for G6PD deficiency in Africa by resistance to severe malaria. Nature 376, 246–9. doi:10.1038/376246a0

Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J., Ackerman, H.C., Campbell, S.J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R., Lander, E.S., 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature 419, 832–7. doi:10.1038/nature01140

Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R., Schaffner, S.F., Lander, E.S., Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., Pasternak, S., Wheeler, D.A., Willis, T.D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, H., Zhou, J., Gabriel, S.B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Onofrio, R.C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Altshuler, D., Shen, Y., Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Sun, W., Wang, H., Wang, Y., Xiong, X., Xu, L., Waye, M.M., Tsui, S.K., Xue, H., Wong, J.T., Galver, L.M., Fan, J.B., Gunderson, K., Murray, S.S., Oliphant, A.R., Chee, M.S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J.F., Phillips, M.S., Roumy, S., Sallee, C., Verner, A., Hudson, T.J., Kwok, P.Y., Cai, D., Koboldt, D.C., Miller, R.D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L.C., Mak, W., Song, Y.Q., Tam, P.K., Nakamura, Y., Kawaguchi, T., Kitamoto, T., Morizono, T., Nagashima, A., Ohnishi, Y., Sekine, A., Tanaka, T., Tsunoda, T., Deloukas, P., Bird, C.P., Delgado, M., Dermitzakis, E.T., Gwilliam, R., Hunt, S., Morrison, J., Powell, D., Stranger, B.E., Whittaker, P., Bentley, D.R., Daly, M.J., de Bakker, P.I., Barrett, J., Chretien, Y.R., Maller, J., McCarroll, S., Patterson, N., Pe'er, I., Price, A., Purcell, S., Richter, D.J., Sabeti, P., Saxena, R., Sham, P.C., Stein, L.D., Krishnan, L., Smith, A.V., Tello-Ruiz, M.K., Thorisson, G.A., Chakravarti, A., Chen, P.E., Cutler, D.J., Kashuk, C.S., Lin, S., Abecasis, G.R., Guan, W., Li, Y., Munro, H.M., Qin, Z.S., Thomas, D.J., McVean, G., Auton, A., Bottolo, L., Cardin, N., Eyheramendy, S., Freeman, C., Marchini, J., Myers, S., Spencer, C., Stephens, M., Donnelly, P., Cardon, L.R., Clarke, G., Evans, D.M., Morris, A.P., Weir, B.S., Johnson, T.A., Mullikin, J.C., Sherry, S.T., Feolo, M., Skol, A., Zhang, H., Matsuda, I., Fukushima, Y., Macer, D.R., Suda, E., Rotimi, C.N., Adebamowo, C.A., Ajayi, I., Aniagwu, T., Marshall, P.A., Nkwodimmah, C., Royal,

C.D., Leppert, M.F., Dixon, M., Peiffer, A., Qiu, R., Kent, A., Kato, K., Niikawa, N., Adewole, I.F., Knoppers, B.M., Foster, M.W., Clayton, E.W., Watkin, J., Muzny, D., Nazareth, L., Sodergren, E., Weinstock, G.M., Yakub, I., Birren, B.W., Wilson, R.K., Fulton, L.L., Rogers, J., Burton, J., Carter, N.P., Clee, C.M., Griffiths, M., Jones, M.C., McLay, K., Plumb, R.W., Ross, M.T., Sims, S.K., Willey, D.L., Chen, Z., Han, H., Kang, L., Godbout, M., Wallenburg, J.C., L'Archeveque, P., Bellemare, G., Saeki, K., An, D., Fu, H., Li, Q., Wang, Z., Wang, R., Holden, A.L., Brooks, L.D., McEwen, J.E., Guyer, M.S., Wang, V.O., Peterson, J.L., Shi, M., Spiegel, J., Sung, L.M., Zacharia, L.F., Collins, F.S., Kennedy, K., Jamieson, R., Stewart, J., 2007. Genome-wide detection and characterization of positive selection in human populations. Nature 449, 913–8. doi:10.1038/nature06250

Sabri, A., Grant, A.V., Cosker, K., Azbaoui, S. El, Abid, A., Abderrahmani Rhorfi, I., Souhi, H., Janah, H., Alaoui-Tahiri, K., Gharbaoui, Y., Benkirane, M., Orlova, M., Boland, A., Deswarte, C., Migaud, M., Bustamante, J., Schurr, E., Boisson-Dupuis, S., Casanova, J.-L., Abel, L., Baghdadi, J. El, 2014. Association Study of Genes Controlling IL-12-dependent IFN-γ Immunity: STAT4 Alleles Increase Risk of Pulmonary Tuberculosis in Morocco. J Infect Dis 210, 611–618. doi:10.1093/infdis/jiu140

Saito, T., Hirai, R., Loo, Y.-M., Owen, D., Johnson, C.L., Sinha, S.C., Akira, S., Fujita, T., Gale, M., 2007. Regulation of innate antiviral defenses through a shared repressor domain in RIG-I and LGP2. Proc. Natl. Acad. Sci. U.S.A. 104, 582–587. doi:10.1073/pnas.0606699104

Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Pääbo, S., Patterson, N., Reich, D., 2014. The genomic landscape of Neanderthal ancestry in present-day humans. Nature 507, 354–357. doi:10.1038/nature12961

Sankararaman, S., Patterson, N., Li, H., Pääbo, S., Reich, D., 2012. The date of interbreeding between Neandertals and modern humans. PLoS Genet. 8, e1002947. doi:10.1371/journal.pgen.1002947

Santos, I.K., Costa, C.H., Krieger, H., Feitosa, M.F., Zurakowski, D., Fardin, B., Gomes, R.B., Weiner, D.L., Harn, D.A., Ezekowitz, R.A., Epstein, J.E., 2001. Mannan-binding lectin enhances susceptibility to visceral leishmaniasis. Infect. Immun. 69, 5212–5215. doi:10.1128/IAI.69.8.5212-5215.2001

Sawcer, S., Ban, M., Maranian, M., Yeo, T.W., Compston, A., Kirby, A., Daly, M.J., De Jager, P.L., Walsh, E., Lander, E.S., Rioux, J.D., Hafler, D.A., Ivinson, A., Rimmler, J., Gregory, S.G., Schmidt, S., Pericak-Vance, M.A., Akesson, E., Hillert, J., Datta, P., Oturai, A., Ryder, L.P., Harbo, H.F., Spurkland, A., Myhr, K.M., Laaksonen, M., Booth, D., Heard, R., Stewart, G., Lincoln, R., Barcellos, L.F., Hauser, S.L., Oksenberg, J.R., Kenealy, S.J., Haines, J.L., 2005. A high-density screen for linkage in multiple sclerosis. American journal of human genetics 77, 454–67. doi:10.1086/444547

Scanga, C.A., Mohan, V.P., Yu, K., Joseph, H., Tanaka, K., Chan, J., Flynn, J.L., 2000. Depletion of CD4(+) T cells causes reactivation of murine persistent tuberculosis despite continued expression of interferon gamma and nitric oxide synthase 2. J. Exp. Med. 192, 347–358.

Schaffner, S.F., Foo, C., Gabriel, S., Reich, D., Daly, M.J., Altshuler, D., 2005. Calibrating a coalescent simulation of human genome sequence variation. Genome Res. 15, 1576–1583. doi:10.1101/gr.3709305

Schattgen, S.A., Fitzgerald, K.A., 2011. The PYHIN protein family as mediators of host defenses. Immunol. Rev. 243, 109–118. doi:10.1111/j.1600-065X.2011.01053.x

Schaub, M.A., Boyle, A.P., Kundaje, A., Batzoglou, S., Snyder, M., 2012. Linking disease associations with regulatory information in the human genome. Genome Res. 22, 1748–1759. doi:10.1101/gr.136127.111

Scheinfeldt, L.B., Tishkoff, S.A., 2013. Recent human adaptation: genomic approaches, interpretation and insights. Nat Rev Genet 14, 692–702. doi:10.1038/nrg3604

Schnall-Levin, M., Zhao, Y., Perrimon, N., Berger, B., 2010. Conserved microRNA targeting in Drosophila is as widespread in coding regions as in 3'UTRs. Proc. Natl. Acad. Sci. U.S.A. 107, 15751–15756. doi:10.1073/pnas.1006172107

Schoggins, J.W., MacDuff, D.A., Imanaka, N., Gainey, M.D., Shrestha, B., Eitson, J.L., Mar, K.B., Richardson, R.B., Ratushny, A.V., Litvak, V., Dabelic, R., Manicassamy, B., Aitchison, J.D., Aderem, A., Elliott, R.M., García-Sastre, A., Racaniello, V., Snijder, E.J., Yokoyama, W.M., Diamond, M.S., Virgin, H.W., Rice, C.M., 2014. Pan-viral specificity of IFN-induced genes reveals new roles for cGAS in innate immunity. Nature 505, 691–695. doi:10.1038/nature12862

Scholes, G., Stein, G., Weiss, J., 1949. Action of X-rays on nucleic acids. Nature 164, 709.

Schroder, K., Tschopp, J., 2010. The inflammasomes. Cell 140, 821–832. doi:10.1016/j.cell.2010.01.040

Schröder, M., Baran, M., Bowie, A.G., 2008. Viral targeting of DEAD box protein 3 reveals its role in TBK1/IKKepsilon-mediated IRF activation. EMBO J. 27, 2147–2157. doi:10.1038/emboj.2008.143

Schwarz, D.S., Hutvágner, G., Du, T., Xu, Z., Aronin, N., Zamore, P.D., 2003. Asymmetry in the assembly of the RNAi enzyme complex. Cell 115, 199–208.

Seguin-Orlando, A., Korneliussen, T.S., Sikora, M., Malaspinas, A.-S., Manica, A., Moltke, I., Albrechtsen, A., Ko, A., Margaryan, A., Moiseyev, V., Goebel, T., Westaway, M., Lambert, D., Khartanovich, V., Wall, J.D., Nigst, P.R., Foley, R.A., Lahr, M.M., Nielsen, R., Orlando, L., Willerslev, E., 2014. Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. Science 346, 1113–1118. doi:10.1126/science.aaa0114

Selbach, M., Schwanhäusser, B., Thierfelder, N., Fang, Z., Khanin, R., Rajewsky, N., 2008. Widespread changes in protein synthesis induced by microRNAs. Nature 455, 58–63. doi:10.1038/nature07228

Sharbati, J., Lewin, A., Kutz-Lohroff, B., Kamal, E., Einspanier, R., Sharbati, S., 2011. Integrated microRNA-mRNA-analysis of human monocyte derived macrophages upon Mycobacterium avium subsp. hominissuis infection. PLoS ONE 6, e20258. doi:10.1371/journal.pone.0020258

Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., Sirotkin, K., 2001. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 29, 308–311.

Shriver, M.D., Kennedy, G.C., Parra, E.J., Lawson, H.A., Sonpar, V., Huang, J., Akey, J.M., Jones, K.W., 2004. The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. Hum. Genomics 1, 274–286.

Siddle, K.J., Tailleux, L., Deschamps, M., Loh, Y.-H.E., Deluen, C., Gicquel, B., Antoniewski, C., Barreiro, L.B., Farinelli, L., Quintana-Murci, L., 2015. Bacterial infection drives the expression dynamics of microRNAs and their isomiRs. PLoS Genet. 11, e1005064. doi:10.1371/journal.pgen.1005064

Sirugo, G., Predazzi, I.M., Bartlett, J., Tacconelli, A., Walther, M., Williams, S.M., 2014. G6PD A- deficiency and severe malaria in The Gambia: heterozygote advantage and possible homozygote disadvantage. The American journal of tropical medicine and hygiene 90, 856–9. doi:10.4269/ajtmh.13-0622

Slack, F.J., Basson, M., Liu, Z., Ambros, V., Horvitz, H.R., Ruvkun, G., 2000. The lin-41 RBCC gene acts in the C. elegans heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor. Mol. Cell 5, 659–669.

Smirnova, I., Hamblin, M.T., McBride, C., Beutler, B., Di Rienzo, A., 2001. Excess of rare amino acid polymorphisms in the Toll-like receptor 4 in humans. Genetics 158, 1657–1664.

Smirnov, D.A., Morley, M., Shin, E., Spielman, R.S., Cheung, V.G., 2009. Genetic analysis of radiation-induced changes in human gene expression. Nature 459, 587–591. doi:10.1038/nature07940

Smith, E.N., Kruglyak, L., 2008. Gene-environment interaction in yeast gene expression. PLoS Biol. 6, e83. doi:10.1371/journal.pbio.0060083

Song, L., Zhang, Z., Grasfeder, L.L., Boyle, A.P., Giresi, P.G., Lee, B.-K., Sheffield, N.C., Gräf, S., Huss, M., Keefe, D., Liu, Z., London, D., McDaniell, R.M., Shibata, Y., Showers, K.A., Simon, J.M., Vales, T., Wang, T., Winter, D., Zhang, Z., Clarke, N.D., Birney, E., Iyer, V.R., Crawford, G.E., Lieb, J.D., Furey, T.S., 2011. Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. Genome Res. 21, 1757–1767. doi:10.1101/gr.121541.111

Soto-Rifo, R., Ohlmann, T., 2013. The role of the DEAD-box RNA helicase DDX3 in mRNA metabolism. Wiley Interdiscip Rev RNA 4, 369–385. doi:10.1002/wrna.1165

Sousa, A.O., Salem, J.I., Lee, F.K., Verçosa, M.C., Cruaud, P., Bloom, B.R., Lagrange, P.H., David, H.L., 1997. An epidemic of tuberculosis with a high rate of tuberculin anergy among a population previously unexposed to tuberculosis, the Yanomami Indians of the Brazilian Amazon. Proc. Natl. Acad. Sci. U.S.A. 94, 13227–13232.

Spielman, R.S., Bastone, L.A., Burdick, J.T., Morley, M., Ewens, W.J., Cheung, V.G., 2007. Common genetic variants account for differences in gene expression among ethnic groups. Nat. Genet. 39, 226–231. doi:10.1038/ng1955

Spinelli, S.V., Diaz, A., D'Attilio, L., Marchesini, M.M., Bogue, C., Bay, M.L., Bottasso, O.A., 2013. Altered microRNA expression levels in mononuclear cells of patients with pulmonary and pleural tuberculosis and their relation with components of the immune response. Mol. Immunol. 53, 265–269. doi:10.1016/j.molimm.2012.08.008

Steimle, V., Otten, L.A., Zufferey, M., Mach, B., 1993. Complementation cloning of an MHC class II transactivator mutated in hereditary MHC class II deficiency (or bare lymphocyte syndrome). Cell 75, 135–146.

Stetson, D.B., Medzhitov, R., 2006. Recognition of cytosolic DNA activates an IRF3-dependent innate immune response. Immunity 24, 93–103. doi:10.1016/j.immuni.2005.12.003

Stranger, B.E., Montgomery, S.B., Dimas, A.S., Parts, L., Stegle, O., Ingle, C.E., Sekowska, M., Smith, G.D., Evans, D., Gutierrez-Arcelus, M., Price, A., Raj, T., Nisbett, J., Nica, A.C., Beazley, C., Durbin, R., Deloukas, P., Dermitzakis, E.T., 2012. Patterns of Cis Regulatory Variation in Diverse Human Populations. PLoS Genet 8, e1002639. doi:10.1371/journal.pgen.1002639

Stranger, B.E., Nica, A.C., Forrest, M.S., Dimas, A., Bird, C.P., Beazley, C., Ingle, C.E., Dunning, M., Flicek, P., Koller, D., Montgomery, S., Tavaré, S., Deloukas, P., Dermitzakis, E.T., 2007. Population genomics of human gene expression. Nat. Genet. 39, 1217–1224. doi:10.1038/ng2142

Sun, L., Wu, J., Du, F., Chen, X., Chen, Z.J., 2013. Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. Science 339, 786–791. doi:10.1126/science.1232458

Tabarsi, P., Marjani, M., Mansouri, N., Farnia, P., Boisson-Dupuis, S., Bustamante, J., Abel, L., Adimi, P., Casanova, J.-L., Mansouri, D., 2011. Lethal tuberculosis in a previously healthy adult with IL-12 receptor deficiency. J. Clin. Immunol. 31, 537–539. doi:10.1007/s10875-011-9523-9

Taganov, K.D., Boldin, M.P., Chang, K.-J., Baltimore, D., 2006. NF-κB-dependent induction of microRNA miR-146, an inhibitor targeted to signaling proteins of innate immune responses. PNAS 103, 12481–12486. doi:10.1073/pnas.0605298103

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585–595.

Tajima, F., 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics 105, 437–460.

Takeda, K., Kaisho, T., Akira, S., 2003. Toll-like receptors. Annu. Rev. Immunol. 21, 335–376. doi:10.1146/annurev.immunol.21.120601.141126

Takeuchi, O., Kawai, T., Sanjo, H., Copeland, N.G., Gilbert, D.J., Jenkins, N.A., Takeda, K., Akira, S., 1999. TLR6: A novel member of an expanding toll-like receptor family. Gene 231, 59–65.

Tanaka, Y., Chen, Z.J., 2012. STING specifies IRF3 phosphorylation by TBK1 in the cytosolic DNA signaling pathway. Sci Signal 5, ra20. doi:10.1126/scisignal.2002521

Taylor, B., Greenstein, J.P., Hollaender, A., 1947. Effects of X-Radiation on Thymus Nucleic Acid. Science 105, 263–264. doi:10.1126/science.105.2723.263

Taylor, S.M., Parobek, C.M., Fairhurst, R.M., 2012. Haemoglobinopathies and the clinical epidemiology of malaria: a systematic review and meta-analysis. The Lancet. Infectious diseases 12, 457–68. doi:10.1016/S1473-3099(12)70055-5

Temme, S., Zacharias, M., Neumann, J., Wohlfromm, S., König, A., Temme, N., Springer, S., Trowsdale, J., Koch, N., 2014. A novel family of human leukocyte antigen class II receptors may have its origin in archaic human species. J. Biol. Chem. 289, 639–653. doi:10.1074/jbc.M113.515767

Teshima, K.M., Coop, G., Przeworski, M., 2006. How reliable are empirical genomic scans for selective sweeps? Genome Res. 16, 702–712. doi:10.1101/gr.5105206

Thalmann, O., Fischer, A., Lankester, F., Pääbo, S., Vigilant, L., 2007. The complex evolutionary history of gorillas: insights from genomic data. Mol. Biol. Evol. 24, 146–158. doi:10.1093/molbev/msl160

Thangaraj, K., Chaubey, G., Kivisild, T., Reddy, A.G., Singh, V.K., Rasalkar, A.A., Singh, L., 2005. Reconstructing the origin of Andaman Islanders. Science 308, 996. doi:10.1126/science.1109987

The International HapMap Consortium, -, -, Gibbs, R.A., Belmont, J.W., Hardenbol, P., Willis, T.D., Yu, F., Yang, H., Ch'ang, L.-Y., Huang, W., Liu, B., Shen, Y., Tam, P.K.-H., Tsui, L.-C., Waye, M.M.Y., Wong, J.T.-F., Zeng, C., Zhang, Q., Chee, M.S., Galver, L.M., Kruglyak, S., Murray, S.S., Oliphant, A.R., Montpetit, A., Hudson, T.J., Chagnon, F., Ferretti, V., Leboeuf, M., Phillips, M.S., Verner, A., Kwok, P.-Y., Duan, S., Lind, D.L., Miller, R.D., Rice, J.P., Saccone, N.L., Taillon-Miller, P., Xiao, M., Nakamura, Y., Sekine, A., Sorimachi, K., Tanaka, T., Tanaka, Y., Tsunoda, T., Yoshino, E., Bentley, D.R., Deloukas, P., Hunt, S., Powell, D., Altshuler, D., Gabriel, S.B., Zhang, H., Zeng, C., Matsuda, I., Fukushima, Y., Macer, D.R., Suda, E., Rotimi, C.N., Adebamowo, C.A., Aniagwu, T., Marshall, P.A., Matthew, O., Nkwodimmah, C., Royal, C.D.M., Leppert, M.F., Dixon, M., Stein, L.D., Cunningham, F., Kanani, A., Thorisson, G.A., Chakravarti, A., Chen, P.E., Cutler, D.J., Kashuk, C.S., Donnelly, P., Marchini, J., McVean, G.A.T., Myers, S.R., Cardon, L.R., Abecasis, G.R., Morris, A., Weir, B.S., Mullikin, J.C., Sherry, S.T., Feolo, M., Altshuler, D., Daly, M.J., Schaffner, S.F., Qiu, R., Kent, A., Dunston, G.M., Kato, K., Niikawa, N., Knoppers, B.M., Foster, M.W., Clayton, E.W., Wang, V.O., Watkin, J., Gibbs, R.A., Belmont, J.W., Sodergren, E., Weinstock, G.M., Wilson, R.K., Fulton, L.L., Rogers, J., Birren, B.W., Han, H., Wang, H., Godbout, M., Wallenburg, J.C., L'Archevêque, P., Bellemare, G., Todani, K., Fujita, T., Tanaka, S., Holden, A.L., Lai, E.H., Collins, F.S., Brooks, L.D., McEwen, J.E., Guyer, M.S., Jordan, E., Peterson, J.L., Spiegel, J., Sung, L.M., Zacharia, L.F., Kennedy, K., Dunn, M.G., Seabrook, R., Shillito, M., Skene, B., Stewart, J.G., Valle (chair), D.L., Clayton (co-chair), E.W., Jorde (co-chair), L.B., Belmont, J.W., Chakravarti, A., Cho, M.K., Duster, T., Foster, M.W., Jasperse, M., Knoppers, B.M., Kwok, P.-Y., Licinio, J., Long, J.C., Marshall, P.A., Ossorio, P.N., Wang, V.O., Rotimi, C.N., Royal, C.D.M., Spallone, P., Terry, S.F., Lander (chair), E.S., Lai (co-chair), E.H., Nickerson (co-chair), D.A., Abecasis, G.R., Altshuler, D., Bentley, D.R., Boehnke, M., Cardon, L.R., Daly, M.J., Deloukas, P., Douglas, J.A., Gabriel, S.B., Hudson, R.R., Hudson, T.J., Kruglyak, L., Kwok, P.-Y., Nakamura, Y., Nussbaum, R.L., Royal, C.D.M., Schaffner, S.F., Sherry, S.T., Stein, L.D., Tanaka, T., -, -, 2003. The International HapMap Project. Nature 426, 789–796. doi:10.1038/nature02168

Thomas, D.L., Thio, C.L., Martin, M.P., Qi, Y., Ge, D., O'Huigin, C., Kidd, J., Kidd, K., Khakoo, S.I., Alexander, G., Goedert, J.J., Kirk, G.D., Donfield, S.M., Rosen, H.R., Tobler, L.H., Busch, M.P., McHutchison, J.G., Goldstein, D.B., Carrington, M., 2009. Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. Nature 461, 798–801. doi:10.1038/nature08463

Thomson, R., Pritchard, J.K., Shen, P., Oefner, P.J., Feldman, M.W., 2000. Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. Proc. Natl. Acad. Sci. U.S.A. 97, 7360–7365.

Thye, T., Nejentsev, S., Intemann, C.D., Browne, E.N., Chinbuah, M.A., Gyapong, J., Osei, I., Owusu-Dabo, E., Zeitels, L.R., Herb, F., Horstmann, R.D., Meyer, C.G., 2009. MCP-1 promoter variant −362C associated with protection from pulmonary tuberculosis in Ghana, West Africa. Hum Mol Genet 18, 381–388. doi:10.1093/hmg/ddn352

Thye, T., Owusu-Dabo, E., Vannberg, F.O., van Crevel, R., Curtis, J., Sahiratmadja, E., Balabanova, Y., Ehmen, C., Muntau, B., Ruge, G., Sievertsen, J., Gyapong, J., Nikolayevskyy, V., Hill, P.C., Sirugo, G., Drobniewski, F., van de Vosse, E., Newport, M., Alisjahbana, B., Nejentsev, S., Ottenhoff, T.H.M., Hill, A.V.S., Horstmann, R.D., Meyer, C.G., 2012. Common variants at 11p13 are associated with susceptibility to tuberculosis. Nat Genet 44, 257–259. doi:10.1038/ng.1080

Thye, T., Vannberg, F.O., Wong, S.H., Owusu-Dabo, E., Osei, I., Gyapong, J., Sirugo, G., Sisay-Joof, F., Enimil, A., Chinbuah, M.A., Floyd, S., Warndorff, D.K., Sichali, L., Malema, S., Crampin, A.C., Ngwira, B., Teo, Y.Y., Small, K., Rockett, K., Kwiatkowski, D., Fine, P.E., Hill, P.C., Newport, M., Lienhardt, C., Adegbola, R.A., Corrah, T., Ziegler, A., African TB Genetics Consortium, Wellcome Trust Case Control Consortium, Morris, A.P., Meyer, C.G., Horstmann, R.D., Hill, A.V.S., 2010. Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. Nat. Genet. 42, 739–741. doi:10.1038/ng.639

Tian, X., Pascal, G., Monget, P., 2009. Evolution and functional divergence of NLRP genes in mammalian reproductive systems. BMC Evol. Biol. 9, 202. doi:10.1186/1471-2148-9-202

Tishkoff, S.A., Reed, F.A., Ranciaro, A., Voight, B.F., Babbitt, C.C., Silverman, J.S., Powell, K., Mortensen, H.M., Hirbo, J.B., Osman, M., Ibrahim, M., Omar, S.A., Lema, G., Nyambo, T.B., Ghori, J., Bumpstead, S., Pritchard, J.K., Wray, G.A., Deloukas, P., 2007. Convergent adaptation of human lactase persistence in Africa and Europe. Nat. Genet. 39, 31–40. doi:10.1038/ng1946

Tishkoff, S.A., Varkonyi, R., Cahinhinan, N., Abbes, S., Argyropoulos, G., Destro-Bisol, G., Drousiotou, A., Dangerfield, B., Lefranc, G., Loiselet, J., Piro, A., Stoneking, M., Tagarelli, A., Tagarelli, G., Touma, E.H., Williams, S.M., Clark, A.G., 2001. Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. Science 293, 455–62. doi:10.1126/science.1061573

Troelsen, J.T., 2005. Adult-type hypolactasia and regulation of lactase expression. Biochim. Biophys. Acta 1723, 19–32. doi:10.1016/j.bbagen.2005.02.003

Troyer, J.L., Nelson, G.W., Lautenberger, J.A., Chinn, L., McIntosh, C., Johnson, R.C., Sezgin, E., Kessing, B., Malasky, M., Hendrickson, S.L., Li, G., Pontius, J., Tang, M., An, P., Winkler, C.A., Limou, S., Le Clerc, S., Delaneau, O., Zagury, J.-F., Schuitemaker, H., van Manen, D., Bream, J.H., Gomperts, E.D., Buchbinder, S., Goedert, J.J., Kirk, G.D., O'Brien, S.J., 2011. Genome-wide association study implicates PARD3B-based AIDS restriction. J. Infect. Dis. 203, 1491–1502. doi:10.1093/infdis/jir046

Turchin, M.C., Chiang, C.W.K., Palmer, C.D., Sankararaman, S., Reich, D., Genetic Investigation of ANthropometric Traits (GIANT) Consortium, Hirschhorn, J.N., 2012. Evidence of widespread selection on standing variation in Europe at height-associated SNPs. Nat. Genet. 44, 1015–1019. doi:10.1038/ng.2368

Underhill, P.A., Kivisild, T., 2007. Use of y chromosome and mitochondrial DNA population structure in tracing human migrations. Annu. Rev. Genet. 41, 539–564. doi:10.1146/annurev.genet.41.110306.130407

Unterholzner, L., Keating, S.E., Baran, M., Horan, K.A., Jensen, S.B., Sharma, S., Sirois, C.M., Jin, T., Latz, E., Xiao, T.S., Fitzgerald, K.A., Paludan, S.R., Bowie, A.G., 2010. IFI16 is an innate immune sensor for intracellular DNA. Nat. Immunol. 11, 997–1004. doi:10.1038/ni.1932

Vabret, N., Blander, J.M., 2013. Sensing Microbial RNA in the Cytosol. Front Immunol 4. doi:10.3389/fimmu.2013.00468

Valadi, H., Ekström, K., Bossios, A., Sjöstrand, M., Lee, J.J., Lötvall, J.O., 2007. Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. Nat Cell Biol 9, 654–659. doi:10.1038/ncb1596

van de Vosse, E., Haverkamp, M.H., Ramirez-Alejo, N., Martinez-Gallo, M., Blancas-Galicia, L., Metin, A., Garty, B.Z., Sun-Tan, Ç., Broides, A., de Paus, R.A., Keskin, Ö., Çağdaş, D., Tezcan, I., Lopez-Ruzafa, E., Aróstegui, J.I., Levy, J., Espinosa-Rosales, F.J., Sanal, Ö., Santos-Argumedo, L., Casanova, J.-L., Boisson-Dupuis, S., van Dissel, J.T., Bustamante, J., 2013. IL-12Rβ1 deficiency: mutation update and description of the IL12RB1 variation database. Hum. Mutat. 34, 1329–1339. doi:10.1002/humu.22380

Vasseur, E., Boniotto, M., Patin, E., Laval, G., Quach, H., Manry, J., Crouau-Roy, B., Quintana-Murci, L., 2012. The evolutionary landscape of cytosolic microbial sensors in humans. Am. J. Hum. Genet. 91, 27–37. doi:10.1016/j.ajhg.2012.05.008

Vasseur, E., Patin, E., Laval, G., Pajon, S., Fornarino, S., Crouau-Roy, B., Quintana-Murci, L., 2011. The selective footprints of viral pressures at the human RIG-I-like receptor family. Hum. Mol. Genet. 20, 4462–4474. doi:10.1093/hmg/ddr377

Vasudevan, S., Tong, Y., Steitz, J.A., 2007. Switching from repression to activation: microRNAs can up-regulate translation. Science 318, 1931–1934. doi:10.1126/science.1149460

Veldhuizen, E.J.A., van Eijk, M., Haagsman, H.P., 2011. The carbohydrate recognition domain of collectins. FEBS J. 278, 3930–3941. doi:10.1111/j.1742-4658.2011.08206.x

Venkataraman, T., Valdes, M., Elsby, R., Kakuta, S., Caceres, G., Saijo, S., Iwakura, Y., Barber, G.N., 2007. Loss of DExD/H box RNA helicase LGP2 manifests disparate antiviral responses. J. Immunol. 178, 6444–6455.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins,

M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M.L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N.N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigó, R., Campbell, M.J., Sjolander, K.V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., Zhu, X., 2001. The sequence of the human genome. Science 291, 1304–1351. doi:10.1126/science.1058040

Verdu, P., Barreiro, L.B., Patin, E., Gessain, A., Cassar, O., Kidd, J.R., Kidd, K.K., Behar, D.M., Froment, A., Heyer, E., Sica, L., Casanova, J.-L., Abel, L., Quintana-Murci, L., 2006. Evolutionary insights into the high worldwide prevalence of MBL2 deficiency alleles. Hum. Mol. Genet. 15, 2650–2658. doi:10.1093/hmg/ddl193

Vernot, B., Akey, J.M., 2015. Complex History of Admixture between Modern Humans and Neandertals. The American Journal of Human Genetics 96, 448–453. doi:10.1016/j.ajhg.2015.01.006

Vernot, B., Stergachis, A.B., Maurano, M.T., Vierstra, J., Neph, S., Thurman, R.E., Stamatoyannopoulos, J.A., Akey, J.M., 2012. Personal and population genomics of human regulatory variation. Genome Res. 22, 1689–1697. doi:10.1101/gr.134890.111

Vitti, J.J., Grossman, S.R., Sabeti, P.C., 2013. Detecting Natural Selection in Genomic Data. Annual Review of Genetics 47, 97–120. doi:10.1146/annurev-genet-111212-133526

Vockley, C.M., Guo, C., Majoros, W.H., Nodzenski, M., Scholtens, D.M., Hayes, M.G., Lowe, W.L., Reddy, T.E., 2015. Massively parallel quantification of the regulatory effects of noncoding genetic variation in a human cohort. Genome Res. doi:10.1101/gr.190090.115

Vogt, G., Bustamante, J., Chapgier, A., Feinberg, J., Boisson Dupuis, S., Picard, C., Mahlaoui, N., Gineau, L., Alcaïs, A., Lamaze, C., Puck, J.M., de Saint Basile, G., Khayat, C.D., Mikhael, R., Casanova, J.-L., 2008. Complementation of a pathogenic IFNGR2 misfolding mutation with modifiers of N-glycosylation. J. Exp. Med. 205, 1729–1737. doi:10.1084/jem.20071987

Vogt, G., Chapgier, A., Yang, K., Chuzhanova, N., Feinberg, J., Fieschi, C., Boisson-Dupuis, S., Alcais, A., Filipe-Santos, O., Bustamante, J., de Beaucoudrey, L., Al-Mohsen, I., Al-Hajjar, S., Al-Ghonaium, A., Adimi, P., Mirsaeidi, M., Khalilzadeh, S., Rosenzweig, S., de la Calle Martin, O., Bauer, T.R., Puck, J.M., Ochs, H.D., Furthner, D., Engelhorn, C., Belohradsky, B., Mansouri, D., Holland, S.M., Schreiber, R.D., Abel, L., Cooper, D.N., Soudais, C., Casanova, J.-L., 2005. Gains of glycosylation comprise an unexpectedly large group of pathogenic mutations. Nat. Genet. 37, 692–700. doi:10.1038/ng1581

Voight, B.F., Kudaravalli, S., Wen, X., Pritchard, J.K., 2006. A map of recent positive selection in the human genome. PLoS Biol. 4, e72. doi:10.1371/journal.pbio.0040072

Wanet, A., Tacheny, A., Arnould, T., Renard, P., 2012. miR-212/132 expression and functions: within and beyond the neuronal compartment. Nucleic Acids Res. 40, 4742–4753. doi:10.1093/nar/gks151

Wang, E.T., Kodama, G., Baldi, P., Moyzis, R.K., 2006. Global landscape of recent inferred Darwinian selection for Homo sapiens. Proc. Natl. Acad. Sci. U.S.A. 103, 135–140. doi:10.1073/pnas.0509691102

Wang, G., 2014. Human antimicrobial peptides and proteins. Pharmaceuticals (Basel) 7, 545–594. doi:10.3390/ph7050545

Wang, Y.-P., Li, K.-B., 2009. Correlation of expression profiles between microRNAs and mRNA targets using NCI-60 data. BMC Genomics 10, 218. doi:10.1186/1471-2164-10-218

Ward, L.D., Kellis, M., 2012. Interpreting noncoding genetic variation in complex traits and human disease. Nat Biotech 30, 1095–1106. doi:10.1038/nbt.2422

Weir, B.S., Hill, W.G., 2002. Estimating F-statistics. Annu. Rev. Genet. 36, 721–750. doi:10.1146/annurev.genet.36.050802.093940

Westra, H.-J., Arends, D., Esko, T., Peters, M.J., Schurmann, C., Schramm, K., Kettunen, J., Yaghootkar, H., Fairfax, B.P., Andiappan, A.K., Li, Y., Fu, J., Karjalainen, J., Platteel, M., Visschedijk, M., Weersma, R.K., Kasela, S., Milani, L., Tserel, L., Peterson, P., Reinmaa, E., Hofman, A., Uitterlinden, A.G., Rivadeneira, F., Homuth, G., Petersmann, A., Lorbeer, R., Prokisch, H., Meitinger, T., Herder, C., Roden, M., Grallert, H., Ripatti, S., Perola, M., Wood, A.R., Melzer, D., Ferrucci, L., Singleton, A.B., Hernandez, D.G., Knight, J.C., Melchiotti, R., Lee, B., Poidinger, M., Zolezzi, F., Larbi, A., Wang, D.Y., van den Berg, L.H., Veldink, J.H., Rotzschke, O., Makino, S., Salomaa, V., Strauch, K., Völker, U., van Meurs, J.B.J., Metspalu, A., Wijmenga, C., Jansen, R.C., Franke, L., 2015. Cell Specific eQTL Analysis without Sorting Cells. PLoS Genet. 11, e1005223. doi:10.1371/journal.pgen.1005223

Westra, H.-J., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M.W., Fairfax, B.P., Schramm, K., Powell, J.E., Zhernakova, A., Zhernakova, D.V., Veldink, J.H., Van den Berg, L.H., Karjalainen, J., Withoff, S., Uitterlinden, A.G., Hofman, A., Rivadeneira, F., Hoen, P.A.C. 't, Reinmaa, E.,

Fischer, K., Nelis, M., Milani, L., Melzer, D., Ferrucci, L., Singleton, A.B., Hernandez, D.G., Nalls, M.A., Homuth, G., Nauck, M., Radke, D., Völker, U., Perola, M., Salomaa, V., Brody, J., Suchy-Dicey, A., Gharib, S.A., Enquobahrie, D.A., Lumley, T., Montgomery, G.W., Makino, S., Prokisch, H., Herder, C., Roden, M., Grallert, H., Meitinger, T., Strauch, K., Li, Y., Jansen, R.C., Visscher, P.M., Knight, J.C., Psaty, B.M., Ripatti, S., Teumer, A., Frayling, T.M., Metspalu, A., van Meurs, J.B.J., Franke, L., 2013. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat. Genet. 45, 1238–1243. doi:10.1038/ng.2756

Wilkinson, R.J., Llewelyn, M., Toossi, Z., Patel, P., Pasvol, G., Lalvani, A., Wright, D., Latif, M., Davidson, R.N., 2000. Influence of vitamin D deficiency and vitamin D receptor polymorphisms on tuberculosis among Gujarati Asians in west London: a case-control study. Lancet 355, 618–621. doi:10.1016/S0140-6736(99)02301-6

Wlasiuk, G., Khan, S., Switzer, W.M., Nachman, M.W., 2009. A history of recurrent positive selection at the toll-like receptor 5 in primates. Mol. Biol. Evol. 26, 937–949. doi:10.1093/molbev/msp018

Wlasiuk, G., Nachman, M.W., 2010. Adaptation and constraint at Toll-like receptors in primates. Mol. Biol. Evol. 27, 2172–2186. doi:10.1093/molbev/msq104

Wong, S.H., Gochhait, S., Malhotra, D., Pettersson, F.H., Teo, Y.Y., Khor, C.C., Rautanen, A., Chapman, S.J., Mills, T.C., Srivastava, A., Rudko, A., Freidin, M.B., Puzyrev, V.P., Ali, S., Aggarwal, S., Chopra, R., Reddy, B.S., Garg, V.K., Roy, S., Meisner, S., Hazra, S.K., Saha, B., Floyd, S., Keating, B.J., Kim, C., Fairfax, B.P., Knight, J.C., Hill, P.C., Adegbola, R.A., Hakonarson, H., Fine, P.E., Pitchappan, R.M., Bamezai, R.N., Hill, A.V., Vannberg, F.O., 2010. Leprosy and the adaptation of human toll-like receptor 1. PLoS pathogens 6, e1000979. doi:10.1371/journal.ppat.1000979

World Health Organization, 2014. Global tuberculosis report. World Health Organization.

World Health Organization, 2013. World Malaria Report 2013. World Health Organization.

Wray, G.A., 2007. The evolutionary significance of cis-regulatory mutations. Nat. Rev. Genet. 8, 206–216. doi:10.1038/nrg2063

Wright, S., 1965. The Interpretation of Population Structure by F-Statistics with Special Regard to Systems of Mating. Evolution 19, 395–420. doi:10.2307/2406450

Wright, S., 1943. Isolation by Distance. Genetics 28, 114–138.

Wright, S., 1931. Evolution in Mendelian Populations. Genetics 16, 97–159.

Wu, J., Lu, C., Diao, N., Zhang, S., Wang, S., Wang, F., Gao, Y., Chen, J., Shao, L., Lu, J., Zhang, X., Weng, X., Wang, H., Zhang, W., Huang, Y., 2012. Analysis of microRNA expression profiling identifies miR-155 and miR-155* as potential diagnostic markers for active tuberculosis: a preliminary study. Hum. Immunol. 73, 31–37. doi:10.1016/j.humimm.2011.10.003

Xu, L., Yang, B.-F., Ai, J., 2013. MicroRNA transport: a new way in cell communication. J. Cell. Physiol. 228, 1713–1719. doi:10.1002/jcp.24344

Yanai, H., Ban, T., Wang, Z., Choi, M.K., Kawamura, T., Negishi, H., Nakasato, M., Lu, Y., Hangai, S., Koshiba, R., Savitsky, D., Ronfani, L., Akira, S., Bianchi, M.E., Honda, K., Tamura, T., Kodama, T., Taniguchi, T., 2009. HMGB proteins function as universal sentinels for nucleic-acid-mediated innate immune responses. Nature 462, 99–103. doi:10.1038/nature08512

Ye, K., Lu, J., Raj, S.M., Gu, Z., 2013. Human expression QTLs are enriched in signals of environmental adaptation. Genome Biol Evol 5, 1689–1701. doi:10.1093/gbe/evt124

Yekta, S., Shih, I.-H., Bartel, D.P., 2004. MicroRNA-directed cleavage of HOXB8 mRNA. Science 304, 594–596. doi:10.1126/science.1097434

Yoneyama, M., Kikuchi, M., Matsumoto, K., Imaizumi, T., Miyagishi, M., Taira, K., Foy, E., Loo, Y.-M., Gale, M., Akira, S., Yonehara, S., Kato, A., Fujita, T., 2005. Shared and unique functions of the DExD/H-box helicases RIG-I, MDA5, and LGP2 in antiviral innate immunity. J. Immunol. 175, 2851–2858.

Yoneyama, M., Kikuchi, M., Natsukawa, T., Shinobu, N., Imaizumi, T., Miyagishi, M., Taira, K., Akira, S., Fujita, T., 2004. The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. Nat. Immunol. 5, 730–737. doi:10.1038/ni1087

Yotova, V., Lefebvre, J.-F., Moreau, C., Gbeha, E., Hovhannesyan, K., Bourgeois, S., Bédarida, S., Azevedo, L., Amorim, A., Sarkisian, T., Avogbe, P.H., Chabi, N., Dicko, M.H., Kou' Santa Amouzou, E.S., Sanni, A., Roberts-Thomson, J., Boettcher, B., Scott, R.J., Labuda, D., 2011. An X-linked haplotype of Neandertal origin is present among all non-African populations. Mol. Biol. Evol. 28, 1957–1962. doi:10.1093/molbev/msr024

Zasloff, M., 2002. Antimicrobial peptides of multicellular organisms. Nature 415, 389–395. doi:10.1038/415389a

Zaykin, D.V., Zhivotovsky, L.A., Czika, W., Shao, S., Wolfinger, R.D., 2007. Combining p-values in large-scale genomics experiments. Pharmaceut. Statist. 6, 217–226. doi:10.1002/pst.304

Zelensky, A.N., Gready, J.E., 2005. The C-type lectin-like domain superfamily. FEBS J. 272, 6179–6217. doi:10.1111/j.1742-4658.2005.05031.x

Zeng, K., Fu, Y.-X., Shi, S., Wu, C.-I., 2006. Statistical Tests for Detecting Positive Selection by Utilizing High-Frequency Variants. Genetics 174, 1431–1439. doi:10.1534/genetics.106.061432

Zhang, F.-R., Huang, W., Chen, S.-M., Sun, L.-D., Liu, H., Li, Y., Cui, Y., Yan, X.-X., Yang, H.-T., Yang, R.-D., Chu, T.-S., Zhang, C., Zhang, L., Han, J.-W., Yu, G.-Q., Quan, C., Yu, Y.-X., Zhang, Z., Shi, B.-Q., Zhang, L.-H., Cheng, H., Wang, C.-Y., Lin, Y., Zheng, H.-F., Fu, X.-A., Zuo, X.-B., Wang, Q., Long, H., Sun, Y.-P., Cheng, Y.-L., Tian, H.-Q., Zhou, F.-S., Liu, H.-X., Lu, W.-S., He, S.-M., Du, W.-L., Shen, M., Jin, Q.-Y., Wang, Y., Low, H.-Q., Erwin, T., Yang, N.-H., Li, J.-Y., Zhao, X., Jiao, Y.-L., Mao, L.-G., Yin, G., Jiang, Z.-X., Wang, X.-D., Yu, J.-P., Hu, Z.-H., Gong, C.-H., Liu, Y.-Q., Liu, R.-Y., Wang, D.-M., Wei, D., Liu, J.-X., Cao, W.-K., Cao, H.-Z., Li, Y.-P., Yan, W.-G., Wei, S.-Y., Wang, K.-J., Hibberd, M.L., Yang, S., Zhang, X.-J., Liu, J.-J., 2009. Genomewide association study of leprosy. N. Engl. J. Med. 361, 2609–2618. doi:10.1056/NEJMoa0903753

Zhang, G., Muglia, L.J., Chakraborty, R., Akey, J.M., Williams, S.M., 2013. Signatures of natural selection on genetic variants affecting complex human traits. Applied & Translational Genomics, SPECIAL ISSUE ON PHARMACOGENOMICS & PERSONALIZED MEDICINE AND SPECIAL ISSUE ON EVOLUTIONARY GENOMICS 2, 78–94. doi:10.1016/j.atg.2013.10.002

Zhang, J., Nielsen, R., Yang, Z., 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol. Biol. Evol. 22, 2472–2479. doi:10.1093/molbev/msi237

Zhang, S.-Y., Boisson-Dupuis, S., Chapgier, A., Yang, K., Bustamante, J., Puel, A., Picard, C., Abel, L., Jouanguy, E., Casanova, J.-L., 2008. Inborn errors of interferon (IFN)-mediated immunity in humans: insights into the respective roles of IFN-α/β, IFN-γ, and IFN-λ in host defense. Immunological Reviews 226, 29–40. doi:10.1111/j.1600-065X.2008.00698.x

Zhong, B., Yang, Y., Li, S., Wang, Y.-Y., Li, Y., Diao, F., Lei, C., He, X., Zhang, L., Tien, P., Shu, H.-B., 2008. The adaptor protein MITA links virus-sensing receptors to IRF3 transcription factor activation. Immunity 29, 538–550. doi:10.1016/j.immuni.2008.09.003

Zietkiewicz, E., Yotova, V., Gehl, D., Wambach, T., Arrieta, I., Batzer, M., Cole, D.E.C., Hechtman, P., Kaplan, F., Modiano, D., Moisan, J.-P., Michalski, R., Labuda, D., 2003. Haplotypes in the dystrophin DNA segment point to a mosaic origin of modern human diversity. Am. J. Hum. Genet. 73, 994–1015. doi:10.1086/378777

Zimmerman, P.A., Ferreira, M.U., Howes, R.E., Mercereau-Puijalon, O., 2013. Red blood cell polymorphism and susceptibility to Plasmodium vivax. Advances in parasitology 81, 27–76. doi:10.1016/B978-0-12-407826-0.00002-3

Zwick, M.E., Cutler, D.J., Chakravarti, A., 2000. Patterns of genetic variation in Mendelian and complex traits. Annual review of genomics and human genetics 1, 387–407. doi:10.1146/annurev.genom.1.1.387

# Annexes

# 1. Travail supplémentaire

PLOS | GENETICS

RESEARCH ARTICLE

# Bacterial Infection Drives the Expression Dynamics of microRNAs and Their isomiRs

**Katherine J. Siddle[1,2☯], Ludovic Tailleux[3☯], Matthieu Deschamps[1,2,4], Yong-Hwee Eddie Loh[1,2], Cécile Deluen[5], Brigitte Gicquel[3], Christophe Antoniewski[6], Luis B. Barreiro[7], Laurent Farinelli[5], Lluís Quintana-Murci[1,2]***

**1** Institut Pasteur, Unit of Human Evolutionary Genetics, Paris, France, **2** Centre National de la Recherche Scientifique, Paris, France, **3** Institut Pasteur, Unit of Mycobacterial Genetics, Paris, France, **4** Université Pierre et Marie Curie, Cellule Pasteur UPMC, Paris, France, **5** Fasteris SA, Plan-les-Ouates, Switzerland, **6** Université Pierre et Marie Curie, Laboratory of Developmental Biology, Paris, France, **7** Ste-Justine Hospital Research Centre and Department of Paediatrics, Faculty of Medicine, University of Montréal, Montréal, Canada

☯ These authors contributed equally to this work.
* quintana@pasteur.fr

## Abstract

The optimal coordination of the transcriptional response of host cells to infection is essential for establishing appropriate immunological outcomes. In this context, the role of microRNAs (miRNAs) – important epigenetic regulators of gene expression – in regulating mammalian immune systems is increasingly well recognised. However, the expression dynamics of miRNAs, and that of their isoforms, in response to infection remains largely unexplored. Here, we characterized the genome-wide miRNA transcriptional responses of human dendritic cells, over time, to various mycobacteria differing in their virulence as well as to other bacteria outside the genus *Mycobacterium*, using small RNA-sequencing. We detected the presence of a core temporal response to infection, shared across bacteria, comprising 49 miRNAs, highlighting a set of miRNAs that may play an essential role in the regulation of basic cellular responses to stress. Despite such broadly shared expression dynamics, we identified specific elements of variation in the miRNA response to infection across bacteria, including a virulence-dependent induction of the miR-132/212 family in response to mycobacterial infections. We also found that infection has a strong impact on both the relative abundance of the miRNA hairpin arms and the expression dynamics of miRNA isoforms. That we observed broadly consistent changes in relative arm expression and isomiR distribution across bacteria suggests that this additional, internal layer of variability in miRNA responses represents an additional source of subtle miRNA-mediated regulation upon infection. Collectively, this study increases our understanding of the dynamism and role of miRNAs in response to bacterial infection, revealing novel features of their internal variability and identifying candidate miRNAs that may contribute to differences in the pathogenicity of mycobacterial infections.

OPEN ACCESS

**Citation:** Siddle KJ, Tailleux L, Deschamps M, Loh Y-HE, Deluen C, Gicquel B, et al. (2015) Bacterial Infection Drives the Expression Dynamics of microRNAs and Their isomiRs. PLoS Genet 11(3): e1005064. doi:10.1371/journal.pgen.1005064

**Editor:** Greg Gibson, Georgia Institute of Technology, UNITED STATES

**Received:** October 8, 2014

**Accepted:** February 9, 2015

**Published:** March 20, 2015



**Data Availability Statement:** The miRNA expression data reported in this manuscript have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under the accession number GSE64142.

**Funding:** The laboratory of LQM has received funding from the Institut Pasteur, the Centre Nationale de la Recherche Scientifique (CNRS), the French Government's Investissement d'Avenir program, Laboratoire d'Excellence "Integrative Biology of Emerging Infectious Diseases" (grant no. ANR-10-LABX-62-IBEID), and from the European Research Council under the European Union's Seventh

Framework Programme (FP/2007–2013)/ERC Grant Agreement No. 281297. The laboratory of LT has received funding from the National Institutes of Health (grant no. NIH AI087658) and the European Commission, within the 7th Framework Programme, (grant no. HEALTH-F3-2009-241745). KJS is a scholar of the Pasteur – Paris University (PPU) International PhD program and was supported by a stipend from the Direction Générale de l'Armement (DGA). YHEL is supported by fellowships from the Fondation pour la Recherche Médicale (FRM) and the Institut Pasteur. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

## Author Summary

MicroRNAs (miRNAs) are small, non-coding RNAs that regulate important cellular processes by inhibiting the expression of gene targets. In recent years, it has become clear that miRNAs play a critical role in the regulation of the immune response to infection, a highly complex phenotype involving the activation of both generic and infection-specific responses. However, it remains unclear to what extent miRNAs are involved in the regulation of these two types of response. Here, focusing on the miRNA response to mycobacteria, pathogens of major public health importance, we present the first comparative, deep sequencing-based analysis of the miRNA response to a panel of bacterial infections. We define a set of miRNAs that play an essential role in basic cellular responses to stress and identify pathogen-specific miRNA responses that reflect mechanisms by which certain pathogens interfere with the host response to infection. In addition, we show that infection can alter the expression level and proportions of miRNA isoforms, transcripts originating from the same miRNA but with slight differences in their nucleotide sequences. This study highlights a novel aspect of miRNA expression dynamics upon infection and increases our understanding of miRNA-mediated mechanisms involved in host cellular responses to infection.

## Introduction

The response of host cells to microbial infection or immune activation is among the most-well studied examples of cellular responses to external stimuli. This response is characterised by marked changes in gene expression [1–6], which require precise coordination to establish appropriate immunological outcomes, ensuring maximal protection against infection while avoiding tissue damage. The crucial role of microRNAs (miRNAs) – small regulatory RNAs that mediate degradation or translational repression of thousands of target mRNAs [7–9] – in regulating mammalian immune systems is increasingly well established. MiRNAs regulate the development and function of immune cells and can have pro-inflammatory or anti-inflammatory effects [10–14]. Furthermore, experimental data indicate that microbial infection alters the miRNA repertoire of host cells [15,16] and that, when aberrantly expressed, miRNAs can contribute to immunity-related pathological conditions, such as infectious or inflammatory diseases, autoimmunity or cancer [12,14,17,18].

The advent of next-generation sequencing, in particular RNA-sequencing, has enabled the exploration of a myriad of novel questions related to miRNA diversity. For example, besides the detection of many novel miRNAs and the description of an increasingly broad array of non-canonical biogenesis pathways producing functional miRNAs [19–21], RNA-seq studies have highlighted the highly dynamic relative abundance of the 5p and 3p arms of the miRNA duplex, a process known as arm-switching [22]. Indeed, following cleavage by Dicer, the miRNA hairpin produces a ~22-bp RNA duplex, one strand of which is preferentially incorporated into the RNA-induced silencing complex (RISC) as a mature, functional miRNA, whereas the other strand has often been thought to be degraded [23]. Previously believed to be static and dictated by the thermodynamic and structural properties of the duplex [24,25], the choice of the dominant miRNA arm has recently been shown to be flexible across species, tissues and developmental stages [22,26–31].

Deep sequencing has also revealed the presence of sequence variation among mature miRNAs – known as isomiRs – shifting the view of miRNAs from single sequences to heterogeneous repertoires of multiple isoforms [32–34]. For a given miRNA, the distribution of

different isomiRs appears to be non-random and can differ between tissues and developmental stages [27,28,30,35–41]. Moreover, that isomiRs appear to act co-operatively with canonical miRNA sequences, targeting common pathways to reduce the signal-to-noise ratio of mRNA targeting [28], suggests that changes in their proportions may have functional implications for gene regulation. Although these studies indicate that arm dominance and isomiR expression are dynamic, the extent to which they can be altered by external stimuli, such as infection, remains unknown.

The characterisation of the host miRNA responses to bacteria has progressed at a slower pace than that of viral and parasitic infections [16], despite the fact that a number of bacteria are responsible for some of the most devastating infectious diseases today. Notable among these is *Mycobacterium tuberculosis* (MTB), the aetiological agent of tuberculosis (TB), the most deadly disease caused by a single bacterial agent [42]. A large number of miRNAs have been recently described as being involved in the response to MTB and other mycobacteria [43–55]. However, the highly heterogeneous nature of these studies—i.e., the use of different mycobacteria, experimental settings (patients, human cells, mice), cell types or tissues, and times post-infection—has precluded any comparison among them, so a clear understanding of the miRNA transcriptional response to MTB is missing. More generally, the extent to which alterations in miRNA expression upon infection are specific to particular pathogens or strains, or instead reflect general responses of host cells to infection, cell activation or inflammation remains to be explored.

Here, we use deep sequencing to characterise the miRNA transcriptional response over time of a key immune cell type – the dendritic cell (DC) – to various mycobacterial strains differing in their virulence as well as to other intracellular bacteria outside the *Mycobacterium* genus. This global, unbiased approach provides a truly comparative picture of the miRNA repertoire, including novel miRNAs, involved in immunity to MTB and, more broadly, bacteria. We defined core bacterial miRNA responses, as well as responses shared between smaller groups of pathogens or detected in a single condition that may reflect particular mechanisms of virulence or suppression. Furthermore, we explored, for the first time, the extent to which infection impacts both the relative abundance of the arms of the miRNA hairpin and the expression dynamics of miRNA isoforms.

## Results

### Expression of annotated and novel miRNAs in dendritic cells

To assess the variability of the genome-wide miRNA response to infection, we exposed human monocyte-derived DCs from healthy donors to a diverse set of bacterial species. This panel included three bacteria of the *Mycobacterium tuberculosis* complex (MTBC), a group of closely related mycobacteria that cause TB in humans or other species. Specifically, we used two virulent strains of *Mycobacterium tuberculosis* – the reference strain H37Rv of the Euro-American lineage (MTB-Rv) and a member of the Beijing strain of the East Asian lineage (MTB-Bj) – as well as the attenuated strain *Mycobacterium bovis* BCG (BCG), widely used as a vaccine against TB. In addition, we included a Gram-positive species, *Staphylococcus epidermidis* (STP), and two Gram-negative bacteria, *Salmonella typhimurium* (SLM) and *Yersinia pseudotuberculosis* (YRS) (S1 Fig.). To study variation in miRNA transcript levels at high resolution, we performed small RNA-sequencing (sRNA-seq) from matched non-infected and infected cells at three time points (4h, 18h and 48h). In total, we generated 1.1 billion reads of 50 bp, corresponding to 116 samples with an average of 9.1 million reads per sample after filtering (GSE64142). Of these, 98% were mappable to the human genome while 85% of reads mapped to miRNAs in miRBase v20 (www.mirbase.org).

Following processing of the data (Methods and S1 Fig.), we detected 387 annotated miRNAs that were present in three or more donors in at least one experimental condition (i.e. cells infected with a given bacterium, or left uninfected, at a given time point) at a depth $\geq$ 50 reads (S1 Table). To identify putative novel miRNAs, we applied a two-step discovery and quantification approach using miRDeep2 [56] (see Methods). Of the 369 predicted hairpins, we detected 18 putative mature miRNAs that were present at a depth $\geq$ 50 reads in three or more donors in at least one experimental condition (S2 Fig.). Of these, 5 corresponded to known snoRNAs, 8 were located in introns, 4 were antisense to genes and 1 was intergenic (S2 Table). Together, this dataset presents a comprehensive, unbiased characterisation of the miRome of steady-state and activated DCs.

## Highly overlapping miRNA responses to diverse bacterial infections

To identify miRNAs whose expression was altered after bacterial infection, we compared infected samples to the corresponding non-infected time point using the package DESeq [57]. A total of 152 miRNAs (38%), of which 145 were previously annotated and 7 were novel, were significantly differentially expressed (FDR-adjusted $p<0.01$ and $|\log_2$ fold change$|>1$) upon encounter with at least one of the six bacteria. For all bacteria, the number of differentially expressed miRNAs increased over time (Fig. 1A and S1 Table). The bacteria that showed the greatest impact on miRNA expression were YRS, MTB-Bj and MTB-Rv ("high-responders"), while BCG, SLM and STP elicited more modest responses ("low-responders") (Fig. 1A). In addition, unsupervised hierarchical clustering clearly separated experimental conditions into three distinct clusters corresponding to the length of infection, with all non-infected conditions clustering with the 4h time point (Fig. 1B), consistent with principle component analysis (S3 Fig.). Overall, these results suggest that the length of infection is a stronger driver of the miRNA response than the identity of the bacterium.

To qualitatively assess the similarity of miRNA responses to the various bacterial infections, we studied the overlap of differentially expressed miRNA sets. We found that over 30% of miRNAs were differentially expressed upon infection with five or more different bacteria, with over 80% shared between at least two independent infections (Fig. 1C). To avoid inflating the dissimilarities between bacteria due to slight differences in fold changes, which were generally highly correlated (S4 Fig.), we also defined miRNAs as significantly differentially expressed when the absolute $\log_2$ fold change was less than 1 if in at least one other experimental condition the change upon infection exceeded this cut-off. Using this threshold, 64% of differentially expressed miRNAs were altered upon infection with at least five of the six bacteria (Fig. 1C). Consistent with their close genetic similarity, MTBC bacteria showed highly overlapping miRNA responses (Fig. 1D). At the same time, less than 10% of miRNAs differentially expressed following MTB-Rv infection were unique to this bacterium, taken as a representative of the MTBC, when compared to more distantly related, non-mycobacterial infections (Fig. 1E). Overall, these results suggest a remarkably consistent miRNA response across diverse bacterial pathogens.

## Temporal miRNA dynamics identifies a core response to bacterial infection

We next investigated whether miRNAs that were differentially expressed upon infection showed similar temporal responses across bacteria, using the Short Time-series Expression Miner (STEM) [58,59]. This program, specifically designed for short time-series datasets, uses the changes in expression observed at each time point to cluster miRNAs according to a set of pre-determined model temporal response profiles. We identified 14 miRNAs that were

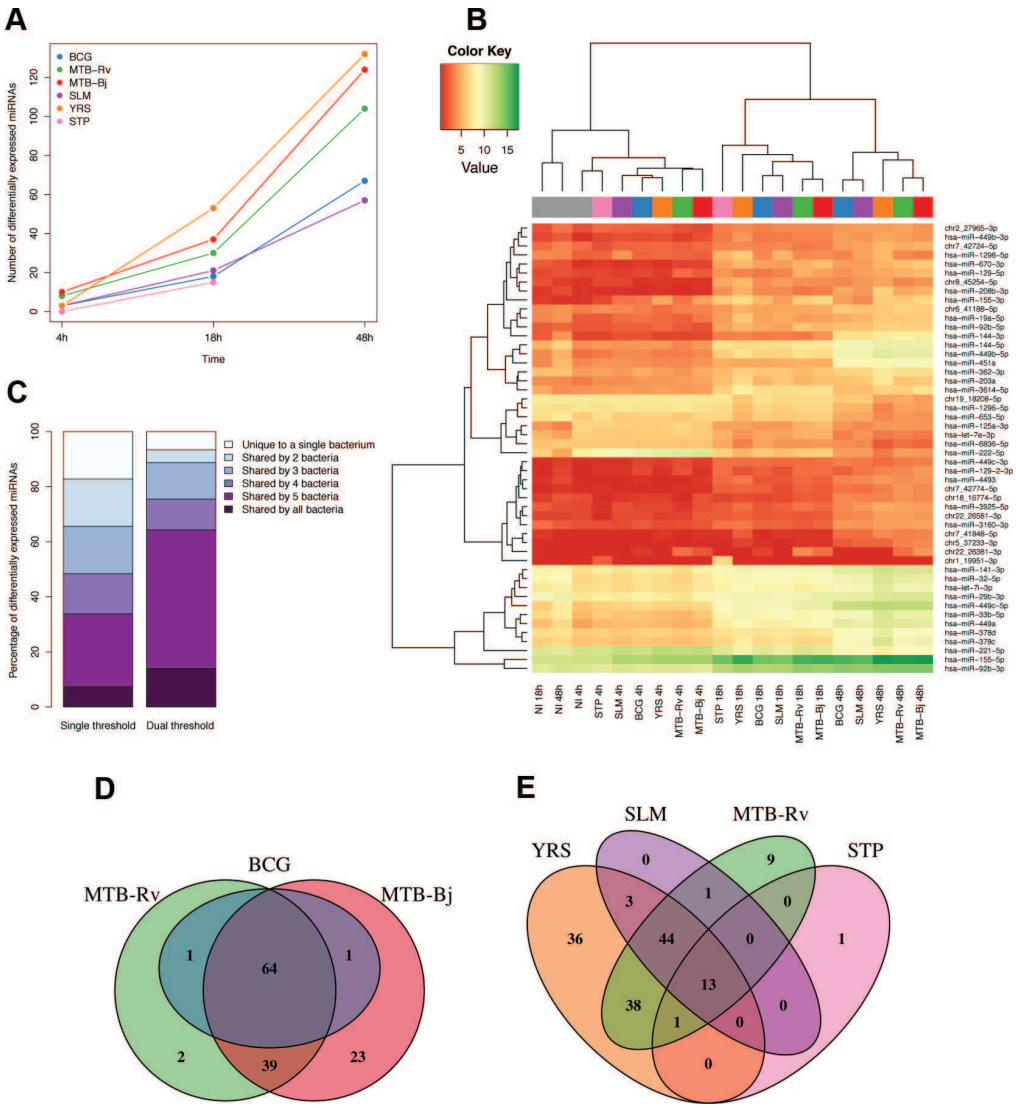



**Fig 1. Differential expression of miRNAs in DCs upon infection with a panel of bacteria.** (A) Numbers of significantly differentially expressed miRNAs upon infection at each time point for each bacterium. As we did not have expression profiles for the 48h time point for STP infection, this point is missing from the plot. (B) Heatmap illustrating the hierarchical clustering of experimental conditions based on the mean expression levels of the 50 most variable miRNAs. (C) Overlap of differentially expressed miRNAs between bacteria using two different significance cut-offs. Left-bar shows overlap using a single cut-off of FDR-adjusted $p<0.01$ and $|\log_2 \text{fold change}|>1$, while the right bar shows the overlap using a secondary cut-off where a miRNA was called as significant if the absolute $\log_2$ fold change was less than 1, if it passed the first more stringent fold-change cut-off upon infection with at least one of the six bacteria. (D and E) Venn diagrams showing the overlap of significantly differentially expressed miRNAs between bacteria of the MTBC (D) and between MTB-Rv and all other non-mycobacterial infections (E).

doi:10.1371/journal.pgen.1005064.g001

assigned to the same model profiles in all bacteria and 35 additional miRNAs that were assigned to highly correlated model profiles in all bacteria (see Methods, Fig. 2A and S3 Table). These 49 miRNAs, which represent the basis of the core miRNA response to bacterial infection in DCs, comprised 27 miRNAs that were upregulated upon infection, 21 that were downregulated and one, miR-222-5p, which was upregulated at 4h but downregulated at later time points. In addition, two novel miRNAs (chr6_41188-5p and chr19_18208-5p) were assigned to highly correlated model profiles in all bacteria (S5 Fig.). To check that these core response

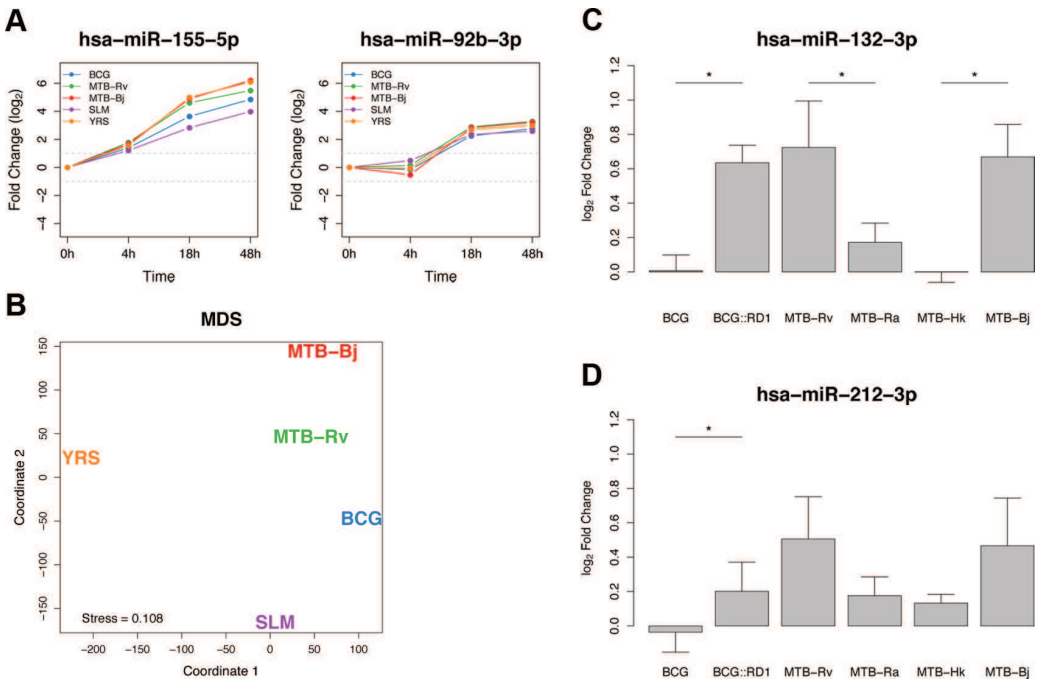



**Fig 2. Shared and specific miRNA responses to bacterial infection.** (A) Plots of the temporal dynamics of two core response miRNAs. As we did not have expression profiles for the 48h time point for STP infection, this bacterium was excluded from the analysis. (B) Multidimensional scaling analysis (MDS) representing the distances between the temporal miRNA responses to different bacterial infections. Distances were based on the sum of edit distances, for each miRNA, between bacteria using STEM-assigned model temporal profiles. (C) The expression of miR-132-3p increased following infection with MTB-Rv and MTB-Bj but not BCG. Transformation of BCG with RD1 from MTB resulted in a significant increase in miR-132-3p expression, not significantly different to that induced by MTB-Rv. Both MTB-Ra and MTB-Hk showed significantly lower miR-132-3p expression than virulent mycobacteria. (D) The induction of miR-212-3p was significantly higher following BCG::RD1 infection, compared to infection with control BCG, and was not significantly different to the response to MTB-Rv. Though the difference between the induction of this miRNA upon infection with virulent or avirulent MTB strains was not significant, the tendencies observed were consistent with miR-132-3p. Significance was calculated using a Mann-Whitney test (* = $p < 0.05$).

doi:10.1371/journal.pgen.1005064.g002

miRNAs indeed showed consistent responses across bacteria, we chose to test two of them – miR-155-5p and miR-92b-3p – by qPCR and confirmed our observations (S6A Fig.).

To provide insight into the impact of this core miRNA response on innate immune cell function, we performed gene ontology category enrichment analysis for gene targets of core miRNAs as predicted by TargetScan [9]. We further restricted our analysis to those targets whose mRNA transcripts have been found to interact with the miRNAs concerned in at least three independent CLIP-seq experiments [60]. We found that a number of relevant biological processes showed some enrichment in high-confidence targets of up-regulated miRNAs, most notably "cellular response to lipopolysaccharide" (S4 Table), while no enrichment was found in high-confidence targets of down-regulated miRNAs. To complement this analysis, we used miRNA and mRNA expression profiles from a previous study in the same cellular system upon MTB-Rv infection [55], to delineate mRNAs whose expression was significantly correlated with that of our core response miRNAs. Interestingly, mRNAs correlated with four core response miRNAs – miR-155–5p, miR-505–3p, miR-7-5p and miR-940 – showed an enrichment in innate immune functions (e.g. innate immune response, immune system process and response to bacterium) [55].

We next used hierarchical clustering to assess the correlation structure among core response miRNAs (Methods). Using the dynamic tree cut algorithm [61], we identified six clusters of highly correlated miRNAs (S3 Table). The two largest clusters contained 20 and 17 miRNAs, respectively. To identify upstream regulators that may explain the coexpression of miRNAs in

these two large clusters, we searched for an enrichment of transcription factor binding in the regions surrounding these miRNAs (Methods). Using transcriptional regulatory relationships from ChIP-seq data obtained from ChIPBase [62], we found that hairpins coding for core response miRNAs belonging to the largest of these clusters (i.e., cluster 1 in S3 Table) were most strongly enriched in the binding of the transcription factor MED12 in their promoter regions ($p = 4\times10^{-3}$). Interestingly, *MED12* has been found to be significantly upregulated upon MTB infection of DCs [63], suggesting a role of this gene in the regulation of this miRNA cluster in response to bacterial infection.

## Distinct miRNA signatures in response to infection by virulent mycobacteria

Despite the generally strong similarity of miRNA responses to all bacteria in our panel (Fig. 1), closer examination revealed a number of more subtle signatures of variability that could be related to differences between bacteria. At the genome-wide level, hierarchical clustering based on miRNA expression levels indicated the presence of sub-clustering at each time point, with some bacteria showing more similar expression profiles following infection (Fig. 1B). This sub-clustering was further supported by a multidimensional scaling analysis (MDS) using the sum of edit distances calculated for each miRNA between bacteria based on their STEM model temporal response profiles (Methods). This showed a strong similarity between the responses to the two virulent MTBC bacteria (MTB-Rv and MTB-Bj), while the response to the attenuated BCG was intermediate between those of MTB-Rv and SLM (Fig. 2B). The separation of YRS in coordinate 1, indicating that the miRNA response to this bacterium was the most distinctive, may be due to the secretion of *Yersinia* outer proteins (Yops), which are unique to this bacterium and have previously been described to modulate host signalling pathways [64].

To further identify miRNA responses that were specific to virulent MTBC (vMTBC) bacteria, we fitted a generalized linear mixed model to test for the effect of infection with a vMTBC strain, while accounting for variability between infection conditions (Methods). We found that the magnitude of the response of 6, 5 and 14 miRNAs, at 4h, 18h and 48h respectively, was specific to infection with MTB-Rv and MTB-Bj (FDR-adjusted $p < 0.01$; S5 Table). This suggests that these miRNAs are part of a virulence-dependent response to mycobacterial infections.

Of the 20 miRNAs that showed a vMTBC-specific response, it is worth noting the presence of miR-132-3p, the only miRNA that was significant at all time points. This miRNA has previously been implicated in the regulation of the inflammatory response [65]. Additionally, both arms of miR-212—the other member of the miR-132/212 family due to their sequence homology and co-localisation on chromosome 17p13.3—also showed a vMTBC-specific response at 18h or 48h. To further investigate the role of virulence in the altered expression of the miR-132/212 family, we studied their response to an extended set of mycobacterial strains that differ in their virulence (Methods). Interestingly, we found that the attenuation and/or inactivation of MTB leads to a significantly lower induction of miR-132–3p, and to a lesser extent miR-212–3p, compared to virulent mycobacteria (Fig. 2C,D and S6 Table). Furthermore, infection with BCG::RD1 – a recombinant strain of BCG containing the RD1 locus, the absence of which accounts, to a large extent, for the attenuation of BCG [66] – significantly increased the induction of miR-132-3p and miR-212-3p, with respect to BCG, attaining a level that was not significantly different from cells infected with the virulent strains (Fig. 2C,D and S6 Table). Overall, these results indicate that the altered expression of the miR-132/miR-212 family is dependent on mycobacterial virulence and, more specifically, that the presence of the virulence-associated RD1 locus is sufficient to account for the stronger induction of the miR-132/212 family among virulent mycobacteria.

## Infection induces changes in the relative expression of the arms of the miRNA hairpin

We next sought to move beyond considering miRNAs as single units to assess the impact of infection on other aspects of miRNA dynamics, from the broader context of the miRNA hairpin to the finer level of internal sequence variability. We first assessed whether infection induces changes in the relative expression of sequences derived from the 5p and 3p arms of the miRNA hairpin. In general, we found that one of the two arms is usually highly dominant. Of the 341 detected mature miRNAs that had a unique genomic alignment, 63% had only one arm expressed at detectable levels, of which one third had no annotated second sequence (Fig. 3A). Of the remaining 27%, 78% had at least a 10-fold dominance of one of the two arms, while only

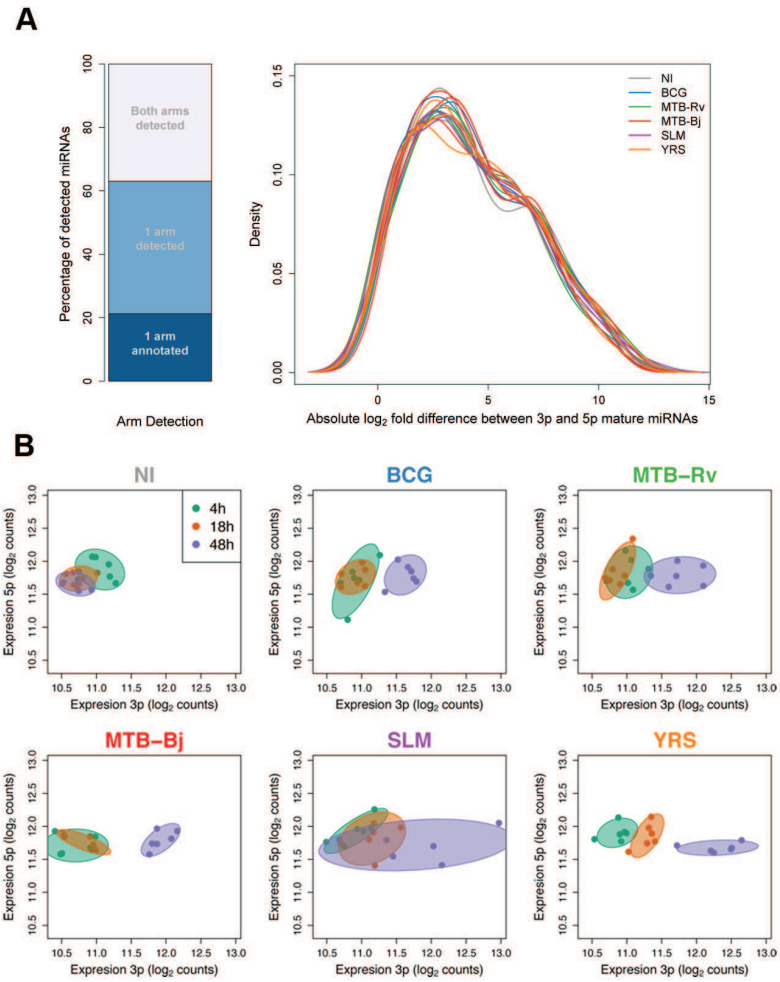



**Fig 3. Changes in relative miRNA arm expression upon infection.** (A) Detection and relative expression of miRNA hairpin arms. Stacked bar plot shows the proportion of detected miRNA hairpins for which either one or both mature miRNA sequences were detected. Density plot shows the distribution of $\log_2$ expression ratios for hairpins where both arms were expressed. (B) An example of one miRNA, miR-361, which showed a strong infection-dependent change in relative arm expression. Axes show the expression of the 3p and 5p arms of all sequenced individuals at each time point. In each panel, colours denote the different time points, and infection conditions are plotted in separate panels. The displacement of points towards the right of the plot in infected samples at later time points shows that the change in the expression ratio of the two arms is due to the increased expression of the 3p arm that, in some conditions, becomes dominant.

doi:10.1371/journal.pgen.1005064.g003

15% showed less than a 2-fold difference between the expression levels of the two arms (Fig. 3A). These figures, however, are likely to underestimate the true dominance as we applied the same expression criteria for the detection of both arms of the hairpin.

We then examined the extent to which infection alters the relative expression of the arms of the miRNA hairpin for the 92 miRNAs for which both arms were detected (Methods). We detected 40 miRNAs that showed a significant change in relative arm expression in at least one experimental condition (FDR-corrected $p < 0.01$), with changes being broadly consistent, in tendency if not in statistical significance, across all bacteria (S7 Table). The majority of these changes grew in magnitude over time. Indeed, 13 hairpins showed differences in relative arm expression between time points in non-infected samples ($p<0.01$), suggesting that time also has a marked, albeit more modest, impact on this feature of miRNA abundance. Of the 27 hairpins (30% of those tested) for which the observed change in the relative expression of the hairpin arms was exclusively associated with infection, three of them (miR-199b, miR-361, miR-582) showed a change in the identity of the dominant miRNA arm upon infection with at least one bacterium (Fig. 3B, S7 Fig. and S7 Table). In most conditions, this change reflected the loss of arm dominance due to a change in the abundance of one of the two arms. However, in two cases (miR-361 and miR-582 expression following YRS infection for 48h) we observed a clear switch in the dominant miRNA sequence. To validate these observations, we measured the expression of one arm-switch miRNA – miR-361 – by qPCR, and obtained a highly concordant tendency of the change in relative expression upon infection (S6B Fig.).

## Abundant changes in isomiR diversity in response to infection

We finally explored the internal sequence diversity of expressed miRNAs (i.e., isomiRs) at the steady-state and upon infection. In general, we detected much greater variability in the end site of isomiRs compared to their start positions (S8A Fig.). Moreover, this variability was dependent on the hairpin arm from which a miRNA was derived, with greater start-site variability among 3p miRNAs and greater end site variability in 5p miRNAs ($p = 3.39\times10^{-14}$ and $p = 3.15\times10^{-20}$, respectively; S8B Fig.). We next classified reads aligning to miRNAs into six groups according to their differences from the canonical miRNA sequence (Fig. 4A). We found that the canonical sequence was the dominant isomiR for only $\sim$50% of miRNAs, with the most common alteration being a change in the end position of the miRNA (3PC), a pattern that was consistently observed across conditions (Fig. 4B and S9 Fig.). Additionally, we observed high isomiR diversity for most miRNAs with the dominant read accounting for only half (median = 49.9%) of all reads aligning to a given miRNA (Fig. 4C).

To assess the impact of infection at the level of individual isomiRs, we searched for differentially expressed isomiRs using DESeq (see Methods) and detected 1,595 isomiRs, corresponding to 235 miRNAs, whose expression was altered upon infection (FDR-corrected p-value $< 0.01$ and $|\log_2$ fold change$| > 1$) (S8 Table). We found a significant overlap between these miRNAs and those that were differentially expressed at the level of total miRNA expression across experimental conditions (p-values = $1.15\times10^{-4}$–$2.03\times10^{-10}$). However, we detected 132 additional miRNAs that show a response to infection at the isomiR level yet are missed when searching only for expression changes at the total miRNA level. Such changes may reflect (i) modest changes in the expression of one or a small number of isoforms that do not have a sufficiently strong cumulative effect on total miRNA levels to be detected, or (ii) changes in the relative proportions of isomiRs upon infection that do not result in an appreciable net gain or loss of reads aligning to the miRNA. When we tested for such changes in the relative expression of isomiRs in response to infection (see Methods), we identified 146 miRNAs that showed a significant change in isomiR proportions upon infection with at least one bacterium (S9

Table). Interestingly, although miRNAs showing changes in isomiR distribution upon infection showed a significant overlap with miRNAs that had one or more significantly differentially expressed isoforms (p-values = $2.48\times10^{-5}$–$7.27\times10^{-10}$), we found only limited overlap between miRNAs showing changes in isomiR distribution and miRNAs that were differentially expressed at the whole miRNA level (p-values = 0.045–1), suggesting that this approach captures an additional aspect of miRNA variation in response to infection.

Changes in isomiR distribution increased over time and were strongly overlapping across bacteria, with only 22% unique to a single bacterial infection (Fig. 4D). For a subset of these

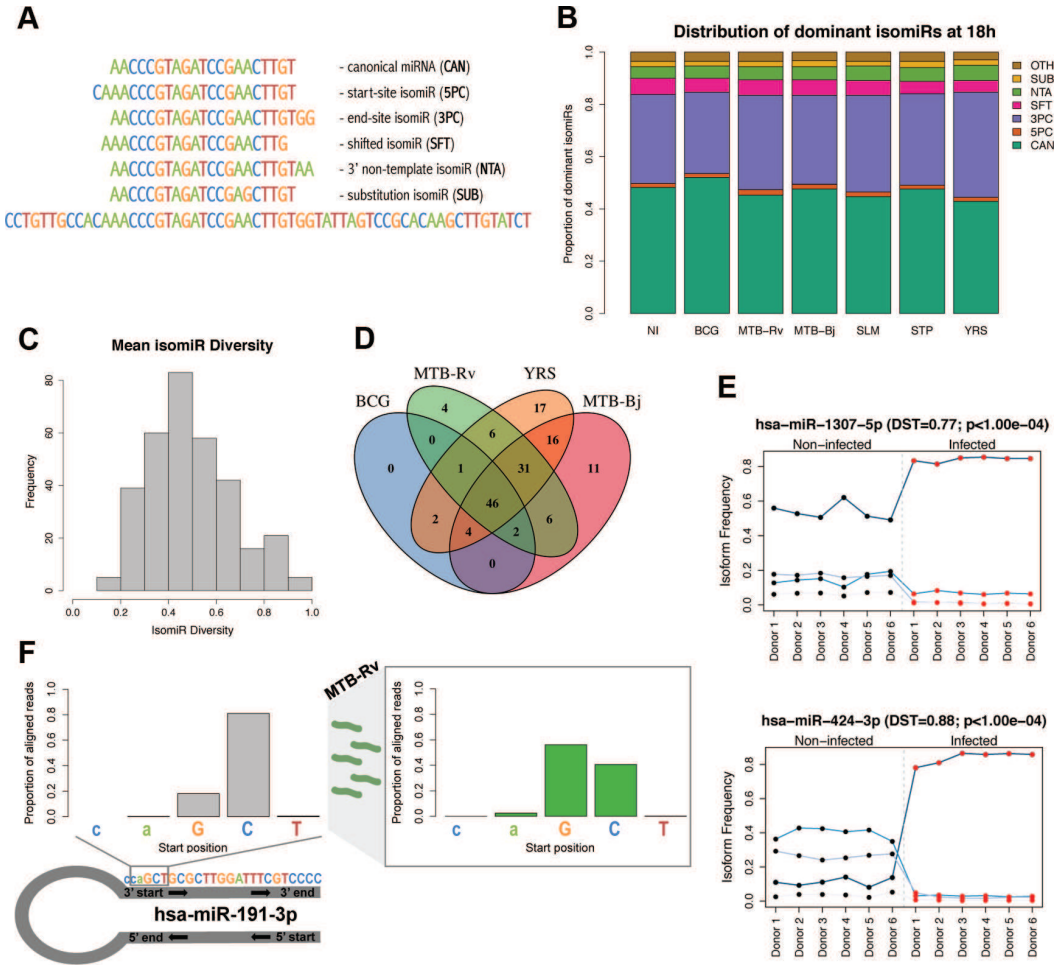



**Fig 4. Dynamic expression of miRNA isomiRs upon infection.** (A) Classification of isomiRs depending on their difference from the annotated mature miRNA sequence (based on [28]). (B) Cumulative barplots showing the proportion of miRNAs for which the dominant sequence was the canonical sequence or, instead, one of six classes of isomiR. Data shown is for 18h post-infection, for other time points see S9 Fig. (C) Histogram showing the mean miRNA diversity (expression of dominant isomiR / total expression of miRNA) across all experimental conditions. (D) Venn diagram showing the overlap of miRNAs showing a significant change in isomiR distribution upon infection with different bacteria. Due to the smaller sample size, we did not perform the analysis upon STP infection. We obtained no significant changes in isomiR distribution following SLM infection, as one individual did not respond to infection (S3 Fig.). (E) Examples of the two characteristic profiles observed for miRNAs showing a change in isomiR distribution upon infection; top: a difference in the proportional expression of each isomiR upon infection, bottom: a change in the dominant isomiR upon infection. Examples are taken from MTB-Rv infection at 48h. Each line represents an individual isomiR and the proportion of the total miRNA expression level accounted for by each isomiR, per donor, is shown before (black dots) and after (red dots) infection. Points are joined together for legibility. (F) The distribution of starting bases of reads aligning to miR-191-3p, which showed a seed shift upon infection with MTB-Rv, MTB-Bj and YRS. Before infection (left), less than 20% of reads start at the first "G" position of the canonical sequence, whereas after infection (right), over 50% of reads start at this position. Proportions of aligned reads shown in the figure are following MTB-Rv infection at 48h.

doi:10.1371/journal.pgen.1005064.g004

miRNAs, this change in isomiR distribution involved a change in the dominant isomiR sequence between non-infected and infected samples (Fig. 4E and S10 Table). For those miRNAs where the canonical miRNA was the dominant sequence in one of the two conditions ( ~ 70% of cases), a switch from the canonical sequence before infection to a non-canonical isoform upon infection was 3× more common than the reverse trend. Consistent with the expression levels of different isomiR classes, the majority of dominant isomiR switches (77–100% per condition) involved a change in the 3' terminus of the most abundant sequence. Interestingly, we identified seven miRNAs – miR-98-3p, miR-140–3p, miR-191-3p, miR-342-5p, miR-548e-5p and miR-2116–3p – that showed a change in the 5' start site of the dominant sequence upon infection with one or more bacteria (Fig. 4F).

To assess the impact of a change in the 5' start site, and hence the functional seed sequence, on potential miRNA targets, we used TargetScan's custom target prediction [67]. We found that 45–87% (mean 71%) of the predicted targets of the annotated miRNA sequence are not predicted targets of the alternative 5' shift isomiR, suggesting that these isomiR changes could profoundly impact upon miRNA-mediated functions. In addition, that miRNAs showing changes in isomiR distribution in response to infection were highly expressed compared to genome-wide miRNA expression levels (Mann-Whitney p-values $8.13\times10^{-4}$–$2.88\times10^{-9}$) suggests that these changes are sufficiently highly abundant to be of functional relevance. These results, together with our observation that infection can also lead to changes in the relative abundance of the hairpin arms, indicate an even greater dynamism in the regulation of miRNA expression than previously appreciated.

## Discussion

In this study, we have shown that the miRNA repertoire involved in the host cellular response to infection is highly similar across a set of different bacteria, both qualitatively – in the identities of differentially expressed miRNAs – and quantitatively – in the high concordance of their expression dynamics upon infection and over time (Fig. 1 and Fig. 2). Specifically, we found that less than 10% of differentially expressed miRNAs are unique to a single bacterium, and defined a set of 49 miRNAs – one third of all differentially expressed miRNAs – that characterises a core response to bacterial infection. Such temporally conserved miRNA responses across bacteria, despite their genetic diversity and differing strategies to manipulate host cell functions [68–70], most likely reflect the activation of common or convergent signalling pathways in response to infection, as has been shown for mRNAs [3,71,72]. For example, the expression of two core response miRNAs—miR-155-5p and let-7i-3p – is known to be induced by the activation of TLRs, key innate immunity receptors that recognise a diverse array of microbial products and the signalling of which is regulated, in part, by miRNAs [73,74].

The consistent changes observed among core miRNAs upon infection raises the question of whether such changes are essential for establishing and maintaining an effective immune response to infection. Though the importance of a small number of miRNAs for the immune response has been described [12,74–76], our understanding of the roles played by miRNAs in the regulation of this and other biological processes remains limited. Some insight can be gained from the study of computationally predicted miRNA targets, however the limited complementarity of animal miRNAs and target sites makes this challenging [7,8]. Additionally, such computational tools are restricted by our current knowledge of the rules of miRNA targeting and do not account for cell-specific interactions, limitations that are reflected in the high false positive and false negative rates of such algorithms [77]. Targets have also been identified through a range of experimental approaches, each of which carry their own limitations [78], and, critically, few targets have been functionally validated. Despite these limitations, the

enrichment analyses of both predicted targets and correlated mRNAs, as well as existing knowledge of the function of some miRNAs, point to the involvement of our set of core response miRNAs in the innate immune response. This suggests that this miRNA response plays a genuine role in the regulation of basic cellular responses to stress, at least in DCs, rather than being a side effect of the immunological changes following infection.

Our results also revealed some important elements of variability in miRNA transcriptional responses between bacteria that may provide insight into bacterial pathogenesis. In this context, we demonstrated that the induction of the miR-132/212 family characterises the response to virulent mycobacteria and is dependent on the presence of the virulence-associated RD1 locus (Fig. 2C,D). Moreover, the reduced induction of miR-132/212 after infection with an attenuated strain that secretes lower levels of the RD1-encoded virulence effector protein ESAT-6 [79] indicates that such an induction is associated with the secretion of this virulence factor. These results raise questions regarding the underlying mechanisms and potential functional consequences of this response, and of the other virulence-dependent miRNA signatures we identified. For example, in light of the role of miR-132 in the negative regulation of the inflammatory response [65,74,80], it is tempting to speculate that the stronger miR-132 induction we observe contributes to the dampening of the early inflammatory response to infection by virulent mycobacteria. Interestingly, reduced early inflammatory responses have been observed in response to modern MTBC lineages, which include MTB-Rv and MTB-Bj, and have been associated with faster progression and increased virulence in macrophages [81]. Further experimental work is now needed to substantiate this hypothesis and, more generally, to identify the specific mycobacterial virulence factors associated with host miRNA expression dynamics.

One of the most interesting findings of our study, made possible by the use of sRNA-seq, is that infection induces changes in both the relative expression of the arms of the miRNA duplex and the distribution of isomiRs (Fig. 3 and Fig. 4). In particular, the induction of strong changes in isomiR distributions, which were highly concordant across individuals and bacteria, highlights the dynamism of miRNA biogenesis and raises important questions regarding the regulation of this process. Several features of our results strongly support that these sequence variants represent true isomiRs showing genuine, infection-dependent changes in their distribution and expression. The isomiRs that we report are expressed at appreciable frequencies and involve more frequent changes at the 3' end of the sequence (S8A Fig.), consistent with the conservation of the seed region [82], which are concordant with known post-transcriptional modifications of miRNAs, such as the non-template addition of, exclusively, "A" and "T" nucleotides [83,84]. In addition, the inverse differences in start and end site variability between miRNAs derived from 3p and 5p arms of the miRNA hairpin (S8B Fig.), supports a greater specificity of Drosha, compared to Dicer, cleavage, as recently suggested [37,41]. More importantly, although specific sequence variants could, theoretically, be the result of errors in the trimming of the sequencing adaptors, these would be expected to occur systematically across conditions and cannot therefore account for the reproducible differences in isomiR expression and/or proportions observed between non-infected and infected cells.

Our results suggest that infection, or the cellular response it elicits, alters one or more of the cellular processes that regulate miRNA expression and isomiR production. The control of miRNA homeostasis is a highly complex and dynamic process involving the transcriptional and post-transcriptional regulation of miRNA expression, biogenesis, loading and decay [19,23]. Even a slight disruption of any of these highly integrated stages could have profound yet variable consequences for miRNA abundance and isomiR diversity as well as, potentially, miRNA functions. For example, isomiR-generating post-transcriptional modifications such as nucleolytic trimming and 3' uridylation and adenylation have been associated with changes in

miRNA stability, loading into the miRISC complex and target gene expression [32,35,83–86], and some viruses have been shown to interfere with these processes [87].

We also found that some miRNAs, including miR-191-3p and miR-342-5p, show a seed shift upon infection. Seed shift isomiRs can have distinct, though overlapping, sets of target genes [28,84,88], as highlighted by our results that show a modest 30% overlap between shifted and canonical sequences, suggesting that the targeting properties of these miRNAs are altered upon infection. More broadly, that the majority of changes at the isomiR level were not detected at the whole miRNA level highlights that focusing only on total miRNA expression misses potentially important changes in miRNA regulation. However, it should be kept in mind that ~85% of all miRNAs had at least one differentially expressed isomiR, emphasising the need to understand further how much of this variability is tolerated by the cell without any biological impact on gene regulation.

In conclusion, our study has reported extensive changes in miRNA expression upon infection that are highly concordant across diverse bacteria and over time. Our results represent the most comprehensive, unbiased view of the similarities in miRNA responses between bacteria to date, and highlight common miRNA-mediated mechanisms that are likely to be essential in the cellular response to stress. Conversely, the detected differences between bacteria may reflect more subtle variations in magnitude and tempo that could, nonetheless, impact on bacterial pathogenesis, such as the case of the induction miR-132-3p. Overall, our findings highlight a novel aspect of miRNA expression dynamics upon infection and increase our understanding of miRNA-mediated mechanisms involved in host cellular responses to infection. In doing so, they provide new perspectives concerning the ways in which infection leads to changes in cellular processes that regulate miRNA expression and isomiR production.

## Materials and Methods

### Ethics statement

Blood samples from nine healthy donors were obtained from the Etablissement Français du Sang. Signed, written consent was obtained from all individuals. The biobank has been declared to and recorded by both the French Ministry of Research and the French Ethics Committee under the reference DC-2008-68 collection 2.

### Bacterial preparation

We infected DCs from six individuals with a panel of six bacteria comprised of: two strains of *Mycobacterium tuberculosis* (H37Rv and GC1237), *Mycobacterium bovis*-BCG Pasteur, *Salmonella typhimurium* strain Keller, *Yersinia pseudotuberculosis* and *Staphylococcus epidermidis* (MTB-Rv, MTB-Bj, BCG, SLM, YRS and STP, respectively). Mycobacteria were grown from a frozen stock to mid-log phase in 7H9 medium supplemented with albumin-dextrose-catalase (Difco). Liquid cultures were grown for up to 12 days and stored at −80°C in 1–2ml aliquots with 10% glycerol. Aliquots were thawed 1 week before infection and bacteria were grown to mid-log phase. Before infection, bacteria were washed 2 times with and re-suspended in 1ml of PBS. Mycobacterial clumps were disassociated by passages through a needle, followed by 5 minutes of sedimentation. Clinical isolates of SLM, STP and YRS were grown on Luria-Bertani agar and stored at −80°C. One day before infection, aliquots were thawed and bacteria grown overnight. 1ml of bacterial culture was grown to mid-log phase shortly prior to infection. Bacterial density in the supernatants was checked at OD600 and confirmed by counting colony-forming units. We infected DCs from three additional individuals with a second panel of six MTBC bacteria comprised of: BCG transformed with the empty cosmid pYUB (BCG) or the same cosmid containing RD1 (BCG::RD1), *Mycobacterium tuberculosis* H37Ra (MTB-Ra), a

non-virulent strain of MTB, heat-killed *Mycobacterium tuberculosis* H37Rv (MTB-Hk), MTB-Rv and MTB-Bj. All strains were prepared as described above for mycobacteria.

## Isolation and infection of DCs

Blood mononuclear cells were isolated by Ficoll-Paque centrifugation. Monocytes were purified from peripheral blood mononuclear cells by positive selection with magnetic CD14 MicroBeads (Miltenyi Biotech). Monocytes were then cultured for 5 days in RPMI-1640 (Invitrogen) supplemented with 10% heat-inactivated FCS (Dutscher), L-glutamine (Invitrogen), GM-CSF (20 ng/mL; Immunotools), and IL-4 (20 ng/mL; Immunotools). Cell cultures were fed every 2 days with complete medium supplemented with the cytokines previously mentioned. The resulting monocyte-derived DCs were infected ( $\sim 2\times10^6$ cells per condition) at an MOI of 1:1 with one of the bacterial panel, or left uninfected, for 1h at 37°C. The cells infected with bacteria of the first panel, or left uninfected, were washed and cultured for a further hour with 50μg/ml gentamycin. The cells were then washed a second time and cultured in complete medium with 5μg/ml gentamycin for an additional 4h, 18h and 48h. In total, we assessed 21 conditions per individual: seven infection conditions (six bacterial infections plus non-infected cells) at three different time points. Due to material limitations and the proliferation rate of the bacteria, we were only able to perform infections and/or recover cells for four of the six individuals and only at 4h and 18h after infection with STP, giving a final total of 116 samples. The cells infected with bacteria of the second panel, or left uninfected, were cultured in complete medium, without gentamycin, for an additional 18h.

## Library preparation and sequencing

Total RNA was extracted using the miRNeasy kit (Qiagen). RNA quantity was assessed using the Qubit (Life Technologies) and RNA quality was assessed using the Agilent 2100 Bioanalyzer with the Nano chip (Agilent Technologies). All samples were of very high quality and showed no signs of degradation (mean RNA integrity number = 9). Sequencing libraries were prepared for each of the 116 samples using the Illumina TruSeq protocol following isolation of low molecular weight (small) RNA fragments. Once prepared, indexed cDNA libraries were pooled (8 or 12 libraries per pool) in equimolar amounts and sequenced with single-end 50bp reads on the Illumina HiSeq2000.

## Pre-processing of raw sequencing reads

Raw reads were first assigned to individual samples based on their multiplexing index, allowing for 1 mismatch. We obtained an average of 11.9 million raw reads per sample with a minimum yield of 6.3 million reads. Next, sequences matching the 3' adaptor sequence were identified and trimmed. A minimum matching of the 6 first bases of the adaptor sequence was required giving reads with final real lengths of 0 to 44 bases. Sequence quality was assessed and subsequent processing performed in R using the Bioconductor package ShortRead [89]. Specifically, we confirmed that base quality (Q) values and per-base GC distributions were within expected ranges and that read length distributions showed an enrichment of reads of the same length as mammalian miRNAs ( $\sim$ 22 bases) (S10 Fig.). We further removed repetitive and low complexity reads. Specifically, we discarded all reads that contained a mononucleotide repeat longer than 10 bases, and those that were >75% mono-, di- or tri-nucleotide repeat or >20% "N" bases. Lastly, we discarded all reads shorter than 16 or longer than 26 bases, corresponding to the length distribution of mammalian miRNAs. After these filtering steps, we obtained an average of 9.1 million (minimum 3.8 million) clean, short reads per sample that were used for small RNA quantification.

## Sequence alignment

Sequences were aligned to the human reference genome (build GRCh37/hg19) using bowtie (version 0.12.7) [90]. We mapped reads allowing for 2 mismatches (-v 2) and reported all best alignments for reads that mapped equally well to more than one genomic location (-a —best —strata). We suppressed reads with more than 50 possible alignments (-m 50). On average, 98% of reads aligned to the human genome, of which 65% aligned uniquely. As miRNAs are short and tend to occur in families that share highly similar sequences, they are particularly susceptible to spurious multiple alignments, a phenomenon called cross-mapping. Around 35% of reads in the present dataset had more than one best alignment. To avoid cross-mapping artefacts, we used a correction strategy that assigns weights to each of the candidate mapping loci of multiply aligning reads [91]. Weights were calculated based on local expression levels and mismatches in the alignment. Python scripts were obtained from the authors and implemented, without modification, as described in the original manuscript [91].

## Prediction of novel miRNAs

We used the program miRDeep2 [56] to detect putative novel miRNAs in our data using a two-step approach. First, we used the mapper module to map all reads to the human reference genome (build GRCh37/hg19) using bowtie (version 0.12.7) with default miRDeep2 alignment parameters [56]. We ran the module on fastq files from all 116 samples by specifying a config file containing a unique identifier for each sample. We removed all sequences containing a base other than A, C, T, G, U or N, collapsed identical reads and output the pooled dataset in fasta format (mapper.pl –d –e –j –h –m –p). We then used the miRDeep2 module to identify novel and known miRNAs in the pooled set of aligned reads from all 116 samples. All reference files containing either mature or precursor sequences of known miRNAs were from miRBase v20, thus we used the –P flag to specify that miRBase identifiers are in post-v18 "5p" and "3p" format. We considered *Pan troglodytes*, *Pan paniscus*, *Gorilla gorilla* and *Pongo pygmaeus* as related species, as described in the miRDeep2 paper [56]. We validated the miRDeep2 mapping and quantification algorithms by comparing read counts of known miRNAs with our own and confirmed that these were highly correlated. We defined a set of high-confidence putative novel miRNAs using a miRDeep score cut-off of 4 (S2 Fig.). We further removed those predictions that overlapped protein-coding exons, based on Ensembl v75 annotations (www.ensembl.org), as well as those that had a predicted hairpin length less than 45 bases or for which no complementary (star) miRNA was detected.

## Expression analysis of annotated and novel miRNA transcripts

We extracted reads aligning to annotated mature miRNA sequences (miRBase v20) or our high confidence set of putative novel miRNAs with at least 75% overlap using BEDTools [92]. As we had libraries that were sequenced to different depths, we normalised the data to give comparable numbers of reads for each sample. Specifically, we used DESeq (version 1.10.1) to calculate a size factor for each library and divided read counts by this factor [57]. To remove lowly or sporadically expressed miRNAs, we kept only those miRNAs with scaled counts of greater than 50 reads in at least three samples from at least one experimental condition.

We used DESeq (version 1.10.1) to identify differentially expressed miRNAs upon infection by fitting a generalized linear model using a negative binomial distribution [57]. Specifically, for each experimental condition we compared miRNA expression levels between non-infected and infected samples at the same time point. As non-infected and infected samples came from the same six donors (four in the case of STP), we controlled for the paired nature of our data by specifying donor identity in our model. We corrected for multiple testing using a stratified

false discovery approach, taking a per-condition Benjamini and Hochberg FDR-corrected p-value $< 0.01$. We also required an absolute $\log_2$ fold change greater than 1. We performed unsupervised hierarchical clustering on the 50 most variable miRNAs (i.e. those with the highest variance in expression across all samples) using mean miRNA expression levels, after variance stabilisation, with the *heatmap.2* function of the R package gplots.

To test for cases where the miRNA response was specific to virulent MTBC bacteria (MTB-Rv and MTB-Bj), while accounting for variability between conditions, we used the R package glmmADMB [93]. We fit a generalized linear mixed model assuming a negative binomial distribution of miRNA expression:

$$\log(\mu_{ik}) = a + b\mathbf{1}_{Infected_k} + c\mathbf{1}_{MTB_k} + d_k$$

where $\mu_{ik}$ is the mean read count for individual $i$ in experimental condition $k$ (bacterial strain or non-infected), $\mathbf{1}_{Infected_k}$ is a categorical variable indicating the presence of infection in condition $k$ (irrespective of the bacterial strain), $\mathbf{1}_{MTB_k}$ is a categorical variable indicating the presence of infection with a virulent MTBC bacterium in condition $k$, $d_k \sim N(0,\sigma_k^2)$ is a random effect of condition $k$, and where a, b and c are all fixed effects to be estimated from the model. We corrected for multiple testing using a stratified false discovery approach, taking a per-condition Benjamini and Hochberg FDR-corrected p-value $< 0.01$.

We also compared, for annotated miRNAs only, the relative expression level of the 5p and 3p arms of the miRNA duplex between experimental conditions. To do so, we selected only miRNAs where both arms of the hairpin were expressed, had a unique genomic location and did not contain a known polymorphism in their mature sequence (MAF>1% in the European CEU population from the 1,000 Genomes Project [94]). We next calculated, for each sample and miRNA hairpin, the ratio between the expression levels of 5p and 3p mature miRNAs ($\log_2$(5p read count / 3p read count)). We then used a paired, two-sided t-test to test for a significant change in this ratio between non-infected and infected samples at the same time point. We used a cut-off of a per-condition Benjamini and Hochberg FDR-corrected p-value $< 0.01$ to determine significance. To assess the effect of time on the relative expression of 5p and 3p-derived miRNAs, we performed the same analysis comparing, pairwise, non-infected samples at the three time points.

## Expression analysis of isomiRs

We extracted reads aligning to annotated mature miRNA sequences as described above and directly normalised the read counts using the same approach as applied for total miRNA expression levels. We removed all reads aligning to miRNAs with multiple genomic alignments and miRNAs that contained a SNP (MAF>1% in the European CEU population from the 1,000 Genomes Project [94]) in their mature sequence. We considered each unique read as a potential miRNA isoform (isomiR). We classified isoforms into six categories: (i) canonical sequences (according to miRBase v20); (ii) changes in start site; (iii) changes in end site; (iv) shifted sequences; (v) containing a substitution; and (vi) non-templated 3' additions; as well as a seventh mixed group of reads containing multiple types of change. We identified isomiRs that were differentially expressed upon infection using the same filtering criteria, approach and significance thresholds as described for total miRNA expression levels. To assess the impact of infection on the distribution of isoforms for a given miRNA, we defined a statistic ($D_{ST}$) that estimates differentiation in isoform proportions between populations of samples. $D_{ST}$ is analogous to $F_{ST}$ [95,96] and $V_{ST}$ [97], which are also used for detecting population differentiation.

$D_{ST}$ varies between 0 and 1 and is calculated by considering,

$$D_{ST} = \frac{D_{all} - (D_a + D_b)/2}{D_{all}}$$

where $D_a$ is the mean Euclidean distance between isomiR proportions of samples from condition *a*, $D_b$ is the mean distance between samples from condition *b*, and $D_{all}$ is the mean distance between samples across conditions. Distances were calculated using the R function *dist* on the proportions of all detected isomiRs for a given miRNA. We calculated p-values based on 10,000 permutations of isomiR proportions, with replacement, per miRNA. We used a cut-off of an empirical p-value $< 0.001$ to determine significance. We found that 97% of the significant changes identified by $D_{ST}$ were also detected using AMOVA [95], in the same experimental condition and using the same statistical cut-off, confirming that our metric captures relevant changes in isomiR distribution.

## Temporal expression profiles of miRNAs in response to infection

We used the Short Time-series Expression Miner (STEM) to characterise the miRNA responses to each bacterium over time [58,59]. This software assigns observations to a pre-determined set of model temporal response profiles based on the correlation coefficient between observed data (i.e., fold-change at each time point) and model profiles. We used default settings except for the maximum unit change between sequential conditions, which we restricted to 1. To account for inter-individual variability, we simultaneously analysed miRNA expression data for all profiled individuals in a given condition using the "repeat data" option. As we did not measure miRNA expression at 0h, we used the "no normalization / add 0" option to set this value to 0. Fold changes were thus calculated by comparing expression levels before and after infection at the same time-point. To account for the fact that STEM allocates a miRNA to a single model profile, even though it may show a strong correlation with one or more additional profiles, we merged clusters where the model profiles were strongly correlated with each other ($r>0.8$). We defined the core miRNA response on the basis of these merged model profiles.

We further used these, STEM-assigned, model profiles to calculate edit distances between pairs of bacteria for a given miRNA based on the number of steps required to change between their respective model profiles. For example, the edit distance between the profile 0,0,1 and 0,1,2 would be 2 while the distance between 0,-1,-1 and 0,-1,-2 would be 1. We used the sum of these pairwise edit distances to represent the difference between a given pair of bacteria in terms of their miRNA response and visualised this by nonmetric multidimensional scaling using the R function *isoMDS* from the MASS package.

## Enrichment of predicted miRNA targets in Gene Ontology categories

We identified high-confidence predicted miRNA targets by combinbing TargetScan target predictions [9,67,98] and miRNA-protein interaction data based on CLIP-seq using the StarBase database of high-confidence interactions [60]. Gene Ontology (GO) biological processes were downloaded from the website of the Gene Ontology Consortium (http://geneontology.org/). We first checked that core response miRNAs were not significantly different from all other miRNAs with respect to their conservation, GC content, and number of predicted targets. We then calculated the proportion of high-confidence predicted targets of core miRNAs in each GO category. Next, we calculated the same measure for 10,000 randomly resampled size-matched miRNA sets and used this to calculate an enrichment p-value. This p-value reflects

the fraction of random miRNA sets having a greater proportion of predicted targets in a given GO category compared to the test set.

### Analysis of miRNA coexpression

We performed hierarchical clustering on miRNA expression levels using the package wgcna with default settings [99,100]. Specifically, we used the dissimilarity of the Topographical Overlap Matrix and average linkage to cluster the 49 core response miRNAs based on normalised miRNA expression levels across all 116 samples. We then used the dynamic tree cut method to cut the branches of the dendrogram to give clusters of highly correlated miRNAs [61]. We used the ChIP-Base database to identify transcription factors (TFs) that bind to miRNA promoter regions, defined as the region from 5kb upstream to 1kb downstream of the transcription start site of the miRNA [62]. As this database contains results from many different tissues and cell lines, we only considered whether binding had been detected, or not, in the promoter region. To identify TFs that were significantly more frequently bound close to coexpressed miRNAs than expected, we compared the average number of bound factors per miRNA to 10,000 randomly sampled size-matched miRNA sets.

### Real-time quantification of miRNAs

Total RNA was extracted using the miRNeasy kit (Qiagen). To quantify miRNA expression levels, cDNA was synthesized and quantitative real-time PCR (qPCR) performed using the Qiagen miScript PCR system and primers (miScript II RT Kit: 218161; miScript SYBR Green PCR kit: 218073; miR-92b-3p MS00032144; miR-132-3p MS00003458; miR-155–5p MS00031486; miR-212-3p MS00003815; miR-361-5p MS00004032; miR-361-3p MS00009555; U6 MS00033740) in a 7900 Real-time PCR system (Applied Biosystem). Relative miRNA expression levels, normalized to the endogenous control U6, were calculated using the $\Delta\Delta C_t$ method [101].

## Supporting Information

**S1 Fig. General workflow and study design.** (A) Experimental conditions used and (B) small RNA sequencing data analysis workflow.
(TIF)

**S2 Fig. Detection of putative novel miRNAs.** (A) Quality control of miRDeep2 novel miRNA discovery using pooled data from all 116 samples. (B) Example of a high confidence putative novel miRNA identified by our approach.
(TIF)

**S3 Fig. Principal Component (PC) analysis of miRNA expression levels.** PC analysis was performed on the normalised read counts of 405 robustly expressed miRNAs reported in the results section of this paper. Components were calculated using the R function *prcomp*, on centred and scaled data. The similarity of the distribution of PCs to known biological variables and potential technical confounders was calculated using a Kruskal-Wallis rank sum test, with the R function *kruskal.test*. We performed the analysis on all samples together (n = 116) and separately on samples from each of our three time points (4h, 18h and 48h). We found that length of infection accounted for the greatest amount of variance between all samples, consistent with our observation of the increase in the number of differentially expressed miRNAs over time and the results of our clustering analysis (Fig. 1A,B). Consistently, when we repeated PCA on samples from each time point separately we observed that the bacteria with which the samples were infected account for an increasing proportion of the variance with time. Specifically, we

observe no correlation of infection with any of the 10 first PCs at 4h, a significant correlation with PC5 at 18h and a significant correlation with PC1 after 48h. We also found, when considering all time points together, a significant correlation between the individual donors and PCs 3 and 4, suggesting a meaningful amount of interindividual variability in miRNA expression. However, we observed no correlation between the PCs and technical variables such as sequencing run, sequencing lane and index sequence, suggesting that these potential technical confounders do not substantially influence our results.
(TIF)

**S4 Fig. Correlation of $\log_2$ fold changes of miRNAs in response to infection between different bacteria at different time points.** We observed a strong correlation in the fold changes of miRNAs upon infection with our panel of diverse bacteria at 18h and 48h. The correlation was less pronounced at 4h for many comparisons, an observation that may be partially accounted for by the globally smaller fold changes at this early time point.
(TIF)

**S5 Fig. Core temporal responses of novel miRNAs.** Two putative novel miRNAs (chr6_41188-5p and chr19_18208-5p) that showed highly concordant changes in expression profiles over time across all bacteria. As no expression profiles were available for the 48h time point for STP infection, this bacterium was excluded from the analysis.
(TIF)

**S6 Fig. Validation of core miRNA and arm-switch miRNA expression profiles by qPCR.** (A) Response to infection of two core miRNAs (miR-155-5p and miR-92b-3p) measured by qPCR. We assessed their induction upon infection with each of our bacteria compared to non-infected cells at the same time point. We confirmed that these miRNAs were significantly induced upon infection with all bacteria at both 18h and 48h. For miR-155-5p, the fold changes were more pronounced at 48h than at 18h, while for miR-92b-3p fold changes were broadly consistent for each bacterium between the two time points, consistent with the temporal induction profiles observed for each by sRNA-seq (Fig. 2A and S3 Table). (B) Relative expression of the arms of the mir-361 hairpin measured by qPCR. We compared their relative expression before and after infection at 18h and 48h. Consistent with our sRNA-seq results, we observed a change in the ratio of the expression of the two arms upon infection. This change was most pronounced at 48h and resulted in a switch in the dominant arm of the miRNA following YRS infection. It should be noted that the fold changes observed by qPCR are much smaller than those obtained by sRNA-seq. This is most likely due to the different way in which expression is measured and calibrated by qPCR, where higher expression is denoted by a lower CT value and normalised to the expression of a housekeeping gene. Importantly, however, the tendencies we observed by sRNA-seq remain the same. The data presented in both panels represent the mean of a duplicate of qPCR calculated for four independent donors (of the six profiled by sRNA-seq). Expression levels were normalised on RNU6-1. The additional two donors as well as the 4h time point were excluded due to limited sample availability.
(TIF)

**S7 Fig. Relative arm expression of three miRNAs showing a change in the dominant arm of the hairpin upon infection with at leas one bacterium.** Significance, compared to the non-infected condition, was calculated using a paired t-test (** = FDR-adjusted $p < 0.01$).
(TIF)

**S8 Fig. Variability around the start and end site of annotated miRNAs.** (A) Greater variability was observed at the end of the miRNA sequence, relative to the end-point of the canonical

sequence, compared to the start. Specifically, while over 95% of reads aligning to a given miRNA had the same start position as that of the canonical sequence, less than 60% of aligned reads had the same end position as the canonical sequence, with extensions compared to the canonical position more common than shortening. (B) The extent of variability at the start and end sites was differentially distributed between mature miRNAs derived from the 3p and 5p arms of the miRNA hairpin. Specifically, a lower proportion of reads aligning to 3p miRNAs shared the canonical start site compared to reads aligning to 5p miRNAs. Conversely, a lower proportion of reads aligning to 5p miRNAs shared the canonical end site compared to reads aligning to 3p miRNAs. In other words, 3p-derived miRNAs showed relatively greater start-site variability while 5p miRNAs showed relatively greater end-site variability.
(TIF)

**S9 Fig. Proportions of dominant isomiR classes.** Cumulative barplots showing the proportion of miRNAs for which the dominant sequence was the canonical sequence or, instead, one of six classes of isomiR at 4h and 48h. The results for the 18h time point are presented in Fig. 4B, and a schematic representation of the isomiR acronyms given in the legend is presented in Fig. 4A.
(TIF)

**S10 Fig. Quality control of Illumina sequencing data.** (A) Base quality, (B) GC content distribution and (C) insert lengths of raw sequence reads. Values shown are for one representative sequencing run of 18 samples. All samples were systematically randomised for sequencing to avoid technical confounders that could prevent the detection of true differences between experimental conditions.
(TIF)

**S1 Table. List of expressed miRNAs with p-values and fold-changes of their differential expression upon infection compared to non-infected samples**
(XLSX)

**S2 Table. Genomic locations of novel miRNA hairpins described in the study.**
(XLSX)

**S3 Table. STEM-assigned model temporal response profiles and wgcna clusters for core response miRNAs.**
(XLSX)

**S4 Table. Fifty most enriched Gene Ontology biological processes among high-confidence predicted targets of up-regulated core response miRNAs.**
(XLSX)

**S5 Table. List of expressed miRNAs with p-values of the specificity of their response following infection with virulent MTBC bacteria.**
(XLSX)

**S6 Table. Differential expression of the miR-132/212 family following infection with the extended MTBC panel.**
(XLSX)

**S7 Table. miRNAs showing significant changes in relative arm expression upon infection.**
(XLSX)

**S8 Table. List of expressed isomiRs with p-values and fold-changes of their differential expression upon infection compared to non-infected samples.**
(XLSX)

**S9 Table. List of expressed miRNAs with p-values and $D_{ST}$ statistics reflecting infection-induced changes in isomiR distribution.**
(XLSX)

**S10 Table. List of the dominant isomiR in each experimental condition.**
(XLSX)

## Acknowledgments

We thank Roland Brosch for the generous donation of recombinant bacterial strains and Maxime Rotival, Guillaume Laval, Frederik Gwinner and Robin Friedman for helpful discussions.

## Author Contributions

Conceived and designed the experiments: KJS LT LQM. Performed the experiments: KJS LT MD CD LF. Analyzed the data: KJS. Contributed reagents/materials/analysis tools: YHEL BG CA LBB. Wrote the paper: KJS LQM.

## References

1. Chaussabel D, Semnani RT, McDowell MA, Sacks D, Sher A, et al. Unique gene expression profiles of human macrophages and dendritic cells to phylogenetically distinct parasites. Blood 2003; 102: 672–681. PMID: 12663451
2. Huang Q, Liu D, Majewski P, Schulte LC, Korn JM, et al. The plasticity of dendritic cell responses to pathogens and their components. Science 2001; 294: 870–875. PMID: 11679675
3. Chevrier N, Mertins P, Artyomov MN, Shalek AK, Iannacone M, et al. Systematic discovery of TLR signaling components delineates viral-sensing circuits. Cell 2011; 147: 853–867. doi: 10.1016/j.cell.2011.10.022 PMID: 22078882
4. Fairfax BP, Humburg P, Makino S, Naranbhai V, Wong D, et al. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. Science 2014; 343: 1246949. doi: 10.1126/science.1246949 PMID: 24604202
5. Gat-Viks I, Chevrier N, Wilentzik R, Eisenhaure T, Raychowdhury R, et al. Deciphering molecular circuits from genetic variation underlying transcriptional responsiveness to stimuli. Nat Biotechnol 2013; 31: 342–349. doi: 10.1038/nbt.2519 PMID: 23503680
6. Lee MN, Ye C, Villani AC, Raj T, Li W, et al. Common genetic variants modulate pathogen-sensing responses in human dendritic cells. Science 2014; 343: 1246980. doi: 10.1126/science.1246980 PMID: 24604203
7. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 2004; 116: 281–297. PMID: 14744438
8. Huntzinger E, Izaurralde E. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nat Rev Genet 2011; 12: 99–110. doi: 10.1038/nrg2936 PMID: 21245828
9. Friedman RC, Farh KK, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res 2009; 19: 92–105. doi: 10.1101/gr.082701.108 PMID: 18955434
10. Chen CZ, Li L, Lodish HF, Bartel DP. MicroRNAs modulate hematopoietic lineage differentiation. Science 2004; 303: 83–86. PMID: 14657504
11. Johnnidis JB, Harris MH, Wheeler RT, Stehling-Sun S, Lam MH, et al. Regulation of progenitor cell proliferation and granulocyte function by microRNA-223. Nature 2008; 451: 1125–1129. doi: 10.1038/nature06607 PMID: 18278031
12. O'Connell RM, Rao DS, Baltimore D. microRNA regulation of inflammatory responses. Annu Rev Immunol 2012; 30: 295–312. doi: 10.1146/annurev-immunol-020711-075013 PMID: 22224773
13. Lodish HF, Zhou B, Liu G, Chen CZ. Micromanagement of the immune system by microRNAs. Nat Rev Immunol 2008; 8: 120–130. doi: 10.1038/nri2252 PMID: 18204468

14. O'Connell RM, Rao DS, Chaudhuri AA, Baltimore D. Physiological and pathological roles for microRNAs in the immune system. Nat Rev Immunol 2010; 10: 111–122. doi: 10.1038/nri2708 PMID: 20098459
15. Cullen BR. Viruses and microRNAs: RISCy interactions with serious consequences. Genes Dev 2011; 25: 1881–1894. doi: 10.1101/gad.17352611 PMID: 21896651
16. Eulalio A, Schulte L, Vogel J. The mammalian microRNA response to bacterial infections. RNA Biol 2012; 9: 742–750. doi: 10.4161/rna.20018 PMID: 22664920
17. Calin GA, Croce CM. MicroRNA signatures in human cancers. Nat Rev Cancer 2006; 6: 857–866. PMID: 17060945
18. Croce CM. Causes and consequences of microRNA dysregulation in cancer. Nat Rev Genet 2009; 10: 704–714. doi: 10.1038/nrg2634 PMID: 19763153
19. Winter J, Jung S, Keller S, Gregory RI, Diederichs S. Many roads to maturity: microRNA biogenesis pathways and their regulation. Nat Cell Biol 2009; 11: 228–234. doi: 10.1038/ncb0309-228 PMID: 19255566
20. Yang JS, Lai EC. Alternative miRNA biogenesis pathways and the interpretation of core miRNA pathway mutants. Mol Cell 2011; 43: 892–903. doi: 10.1016/j.molcel.2011.07.024 PMID: 21925378
21. Creighton CJ, Reid JG, Gunaratne PH. Expression profiling of microRNAs by deep sequencing. Brief Bioinform 2009; 10: 490–497. doi: 10.1093/bib/bbp019 PMID: 19332473
22. Griffiths-Jones S, Hui JH, Marco A, Ronshaugen M. MicroRNA evolution by arm switching. EMBO Rep 2011; 12: 172–177. doi: 10.1038/embor.2010.191 PMID: 21212805
23. Krol J, Loedige I, Filipowicz W. The widespread regulation of microRNA biogenesis, function and decay. Nat Rev Genet 2010; 11: 597–610. doi: 10.1038/nrg2843 PMID: 20661255
24. Khvorova A, Reynolds A, Jayasena SD. Functional siRNAs and miRNAs exhibit strand bias. Cell 2003; 115: 209–216. PMID: 14567918
25. Schwarz DS, Hutvagner G, Du T, Xu Z, Aronin N, et al. Asymmetry in the assembly of the RNAi enzyme complex. Cell 2003; 115: 199–208. PMID: 14567917
26. Chang HT, Li SC, Ho MR, Pan HW, Ger LP, et al. Comprehensive analysis of microRNAs in breast cancer. BMC Genomics 2012; 13 Suppl 7: S18. doi: 10.1186/1471-2164-13-S7-S18 PMID: 23281739
27. Chiang HR, Schoenfeld LW, Ruby JG, Auyeung VC, Spies N, et al. Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. Genes Dev 2010; 24: 992–1009. doi: 10.1101/gad.1884710 PMID: 20413612
28. Cloonan N, Wani S, Xu Q, Gu J, Lea K, et al. MicroRNAs and their isomiRs function cooperatively to target common biological pathways. Genome Biol 2011; 12: R126. doi: 10.1186/gb-2011-12-12-r126 PMID: 22208850
29. Jagadeeswaran G, Zheng Y, Sumathipala N, Jiang H, Arrese EL, et al. Deep sequencing of small RNA libraries reveals dynamic regulation of conserved and novel microRNAs and microRNA-stars during silkworm development. BMC Genomics 2010; 11: 52. doi: 10.1186/1471-2164-11-52 PMID: 20089182
30. Li SC, Liao YL, Ho MR, Tsai KW, Lai CH, et al. miRNA arm selection and isomiR distribution in gastric cancer. BMC Genomics 2012; 13 Suppl 1: S13. doi: 10.1186/1471-2164-13-S1-S13 PMID: 22369582
31. Marco A, Hui JH, Ronshaugen M, Griffiths-Jones S. Functional shifts in insect microRNA evolution. Genome Biol Evol 2010; 2: 686–696. doi: 10.1093/gbe/evq053 PMID: 20817720
32. Ameres SL, Zamore PD. Diversifying microRNA sequence and function. Nat Rev Mol Cell Biol 2013; 14: 475–488. doi: 10.1038/nrm3611 PMID: 23800994
33. Glazov EA, Cottee PA, Barris WC, Moore RJ, Dalrymple BP, et al. A microRNA catalog of the developing chicken embryo identified by a deep sequencing approach. Genome Res 2008; 18: 957–964. doi: 10.1101/gr.074740.107 PMID: 18469162
34. Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, et al. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res 2008; 18: 610–621. doi: 10.1101/gr.7179508 PMID: 18285502
35. Fernandez-Valverde SL, Taft RJ, Mattick JS. Dynamic isomiR regulation in Drosophila development. RNA 2010; 16: 1881–1888. doi: 10.1261/rna.2379610 PMID: 20805289
36. Lee LW, Zhang S, Etheridge A, Ma L, Martin D, et al. Complexity of the microRNA repertoire revealed by next-generation sequencing. RNA 2010; 16: 2170–2180. doi: 10.1261/rna.2225110 PMID: 20876832
37. Humphreys DT, Hynes CJ, Patel HR, Wei GH, Cannon L, et al. Complexity of murine cardiomyocyte miRNA biogenesis, sequence variant expression and function. PLoS One 2012; 7: e30933. doi: 10.1371/journal.pone.0030933 PMID: 22319597

38. Llorens F, Hummel M, Pantano L, Pastor X, Vivancos A, et al. Microarray and deep sequencing cross-platform analysis of the mirRNome and isomiR variation in response to epidermal growth factor. BMC Genomics 2013; 14: 371. doi: 10.1186/1471-2164-14-371 PMID: 23724959

39. Marti E, Pantano L, Banez-Coronel M, Llorens F, Minones-Moyano E, et al. A myriad of miRNA variants in control and Huntington's disease brain regions detected by massively parallel sequencing. Nucleic Acids Res 2010; 38: 7219–7235. doi: 10.1093/nar/gkq575 PMID: 20591823

40. Vaz C, Ahmad HM, Bharti R, Pandey P, Kumar L, et al. Analysis of the microRNA transcriptome and expression of different isomiRs in human peripheral blood mononuclear cells. BMC Res Notes 2013; 6: 390. doi: 10.1186/1756-0500-6-390 PMID: 24073671

41. Zhou H, Arcila ML, Li Z, Lee EJ, Henzler C, et al. Deep annotation of mouse iso-miR and iso-moR variation. Nucleic Acids Res 2012; 40: 5864–5875. doi: 10.1093/nar/gks247 PMID: 22434881

42. WHO (2013) Global Tuberculosis Report 2013. World Health Organization, Geneva.

43. Fu Y, Yi Z, Wu X, Li J, Xu F. Circulating microRNAs in patients with active pulmonary tuberculosis. J Clin Microbiol 2011; 49: 4246–4251. doi: 10.1128/JCM.05459-11 PMID: 21998423

44. Kumar R, Halder P, Sahu SK, Kumar M, Kumari M, et al. Identification of a novel role of ESAT-6-dependent miR-155 induction during infection of macrophages with Mycobacterium tuberculosis. Cell Microbiol 2012; 14: 1620–1631. doi: 10.1111/j.1462-5822.2012.01827.x PMID: 22712528

45. Liu Y, Wang X, Jiang J, Cao Z, Yang B, et al. Modulation of T cell cytokine production by miR-144* with elevated expression in patients with pulmonary tuberculosis. Mol Immunol 2011; 48: 1084–1090. doi: 10.1016/j.molimm.2011.02.001 PMID: 21367459

46. Ma F, Xu S, Liu X, Zhang Q, Xu X, et al. The microRNA miR-29 controls innate and adaptive immune responses to intracellular bacterial infection by targeting interferon-gamma. Nat Immunol 2011; 12: 861–869. doi: 10.1038/ni.2073 PMID: 21785411

47. Maertzdorf J, Weiner J 3rd, Mollenkopf HJ, Bauer T, Prasse A, et al. Common patterns and disease-related signatures in tuberculosis and sarcoidosis. Proc Natl Acad Sci U S A 2012; 109: 7853–7858. doi: 10.1073/pnas.1121072109 PMID: 22547807

48. Rajaram MV, Ni B, Morris JD, Brooks MN, Carlson TK, et al. Mycobacterium tuberculosis lipomannan blocks TNF biosynthesis by regulating macrophage MAPK-activated protein kinase 2 (MK2) and microRNA miR-125b. Proc Natl Acad Sci U S A 2011; 108: 17408–17413. doi: 10.1073/pnas.1112660108 PMID: 21969554

49. Sharbati J, Lewin A, Kutz-Lohroff B, Kamal E, Einspanier R, et al. Integrated microRNA-mRNA-analysis of human monocyte derived macrophages upon Mycobacterium avium subsp. hominissuis infection. PLoS One 2011; 6: e20258. doi: 10.1371/journal.pone.0020258 PMID: 21629653

50. Singh Y, Kaul V, Mehra A, Chatterjee S, Tousif S, et al. Mycobacterium tuberculosis controls microRNA-99b (miR-99b) expression in infected murine dendritic cells to modulate host immunity. J Biol Chem 2013; 288: 5056–5061. doi: 10.1074/jbc.C112.439778 PMID: 23233675

51. Spinelli SV, Diaz A, D'Attilio L, Marchesini MM, Bogue C, et al. Altered microRNA expression levels in mononuclear cells of patients with pulmonary and pleural tuberculosis and their relation with components of the immune response. Mol Immunol 2013; 53: 265–269. doi: 10.1016/j.molimm.2012.08.008 PMID: 22964481

52. Wang C, Yang S, Sun G, Tang X, Lu S, et al. Comparative miRNA expression profiles in individuals with latent and active tuberculosis. PLoS One 2011; 6: e25832. doi: 10.1371/journal.pone.0025832 PMID: 22003408

53. Wu J, Lu C, Diao N, Zhang S, Wang S, et al. Analysis of microRNA expression profiling identifies miR-155 and miR-155* as potential diagnostic markers for active tuberculosis: a preliminary study. Hum Immunol 2012; 73: 31–37. doi: 10.1016/j.humimm.2011.10.003 PMID: 22037148

54. Yi Z, Fu Y, Ji R, Li R, Guan Z. Altered microRNA signatures in sputum of patients with active pulmonary tuberculosis. PLoS One 2012; 7: e43184. doi: 10.1371/journal.pone.0043184 PMID: 22900099

55. Siddle KJ, Deschamps M, Tailleux L, Nedelec Y, Pothlichet J, et al. A genomic portrait of the genetic architecture and regulatory impact of microRNA expression in response to infection. Genome Res 2014; 24: 850–859. doi: 10.1101/gr.161471.113 PMID: 24482540

56. Friedlander MR, Mackowiak SD, Li N, Chen W, Rajewsky N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res 2012; 40: 37–52. doi: 10.1093/nar/gkr688 PMID: 21911355

57. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol 2010; 11: R106. doi: 10.1186/gb-2010-11-10-r106 PMID: 20979621

58. Ernst J, Bar-Joseph Z. STEM: a tool for the analysis of short time series gene expression data. BMC Bioinformatics 2006; 7: 191. PMID: 16597342

59. Ernst J, Nau GJ, Bar-Joseph Z. Clustering short time series gene expression data. Bioinformatics 2005; 21 Suppl 1: i159–168. PMID: 15961453

60. Yang JH, Li JH, Shao P, Zhou H, Chen YQ, et al. starBase: a database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. Nucleic Acids Res 2011; 39: D202–209. doi: 10.1093/nar/gkq1056 PMID: 21037263

61. Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics 2008; 24: 719–720. PMID: 18024473

62. Yang JH, Li JH, Jiang S, Zhou H, Qu LH. ChIPBase: a database for decoding the transcriptional regulation of long non-coding RNA and microRNA genes from ChIP-Seq data. Nucleic Acids Res 2013; 41: D177–187. doi: 10.1093/nar/gks1060 PMID: 23161675

63. Barreiro LB, Tailleux L, Pai AA, Gicquel B, Marioni JC, et al. Deciphering the genetic architecture of variation in the immune response to Mycobacterium tuberculosis infection. Proc Natl Acad Sci U S A 2012; 109: 1204–1209. doi: 10.1073/pnas.1115761109 PMID: 22233810

64. Viboud GI, Bliska JB. Yersinia outer proteins: role in modulation of host cell signaling responses and pathogenesis. Annu Rev Microbiol 2005; 59: 69–89. PMID: 15847602

65. Nahid MA, Yao B, Dominguez-Gutierrez PR, Kesavalu L, Satoh M, et al. Regulation of TLR2-mediated tolerance and cross-tolerance through IRAK4 modulation by miR-132 and miR-212. J Immunol 2013; 190: 1250–1263. doi: 10.4049/jimmunol.1103060 PMID: 23264652

66. Pym AS, Brodin P, Brosch R, Huerre M, Cole ST. Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines Mycobacterium bovis BCG and Mycobacterium microti. Mol Microbiol 2002; 46: 709–717. PMID: 12410828

67. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 2005; 120: 15–20. PMID: 15652477

68. Bueno MJ, Perez de Castro I, Malumbres M. Control of cell proliferation pathways by microRNAs. Cell Cycle 2008; 7: 3143–3148. PMID: 18843198

69. Galindo CL, Rosenzweig JA, Kirtley ML, Chopra AK. Pathogenesis of Y. enterocolitica and Y. pseudotuberculosis in Human Yersiniosis. J Pathog 2011; 2011: 182051. doi: 10.4061/2011/182051 PMID: 22567322

70. van Kooyk Y, Geijtenbeek TB. DC-SIGN: escape mechanism for pathogens. Nat Rev Immunol 2003; 3: 697–709. PMID: 12949494

71. Jenner RG, Young RA. Insights into host responses against pathogens from transcriptional profiling. Nat Rev Microbiol 2005; 3: 281–294. PMID: 15806094

72. Amit I, Garber M, Chevrier N, Leite AP, Donner Y, et al. Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. Science 2009; 326: 257–263. doi: 10.1126/science.1179050 PMID: 19729616

73. He X, Jing Z, Cheng G. MicroRNAs: new regulators of Toll-like receptor signalling pathways. Biomed Res Int 2014; 2014: 945169. doi: 10.1155/2014/945169 PMID: 24772440

74. O'Neill LA, Sheedy FJ, McCoy CE. MicroRNAs: the fine-tuners of Toll-like receptor signalling. Nat Rev Immunol 2011; 11: 163–175. doi: 10.1038/nri2957 PMID: 21331081

75. Chen CZ, Schaffert S, Fragoso R, Loh C. Regulation of immune responses and tolerance: the microRNA perspective. Immunol Rev 2013; 253: 112–128. doi: 10.1111/imr.12060 PMID: 23550642

76. Li Y, Shi X. MicroRNAs in the regulation of TLR and RIG-I pathways. Cell Mol Immunol 2013; 10: 65–71. doi: 10.1038/cmi.2012.55 PMID: 23262976

77. Alexiou P, Maragkakis M, Papadopoulos GL, Reczko M, Hatzigeorgiou AG. Lost in translation: an assessment and perspective for computational microRNA target identification. Bioinformatics 2009; 25: 3049–3055. doi: 10.1093/bioinformatics/btp565 PMID: 19789267

78. Martinez-Sanchez A, Murphy CL. MicroRNA Target Identification-Experimental Approaches. Biology (Basel) 2013; 2: 189–205. doi: 10.3390/biology2010189 PMID: 24832658

79. Frigui W, Bottai D, Majlessi L, Monot M, Josselin E, et al. Control of M. tuberculosis ESAT-6 secretion and specific T cell recognition by PhoP. PLoS Pathog 2008; 4: e33. doi: 10.1371/journal.ppat.0040033 PMID: 18282096

80. Taganov KD, Boldin MP, Chang KJ, Baltimore D. NF-kappaB-dependent induction of microRNA miR-146, an inhibitor targeted to signaling proteins of innate immune responses. Proc Natl Acad Sci U S A 2006; 103: 12481–12486. PMID: 16885212

81. Portevin D, Gagneux S, Comas I, Young D. Human macrophage responses to clinical isolates from the Mycobacterium tuberculosis complex discriminate between ancient and modern lineages. PLoS Pathog 2011; 7: e1001307. doi: 10.1371/journal.ppat.1001307 PMID: 21408618

82. Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB. Prediction of mammalian microRNA targets. Cell 2003; 115: 787–798. PMID: 14697198

83. Burroughs AM, Ando Y, de Hoon MJ, Tomaru Y, Nishibu T, et al. A comprehensive survey of 3' animal miRNA modification events and a possible role for 3' adenylation in modulating miRNA targeting effectiveness. Genome Res 2010; 20: 1398–1410. doi: 10.1101/gr.106054.110 PMID: 20719920

84. Neilsen CT, Goodall GJ, Bracken CP. IsomiRs—the overlooked repertoire in the dynamic microRNAome. Trends Genet 2012; 28: 544–549. doi: 10.1016/j.tig.2012.07.005 PMID: 22883467

85. Jones MR, Quinton LJ, Blahna MT, Neilson JR, Fu S, et al. Zcchc11-dependent uridylation of microRNA directs cytokine expression. Nat Cell Biol 2009; 11: 1157–1163. doi: 10.1038/ncb1931 PMID: 19701194

86. Katoh T, Sakaguchi Y, Miyauchi K, Suzuki T, Kashiwabara S, et al. Selective stabilization of mammalian microRNAs by 3' adenylation mediated by the cytoplasmic poly(A) polymerase GLD-2. Genes Dev 2009; 23: 433–438. doi: 10.1101/gad.1761509 PMID: 19240131

87. Backes S, Shapiro JS, Sabin LR, Pham AM, Reyes I, et al. Degradation of host microRNAs by poxvirus poly(A) polymerase reveals terminal RNA methylation as a protective antiviral mechanism. Cell Host Microbe 2012; 12: 200–210. doi: 10.1016/j.chom.2012.05.019 PMID: 22901540

88. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell 2009; 136: 215–233. doi: 10.1016/j.cell.2009.01.002 PMID: 19167326

89. Morgan M, Anders S, Lawrence M, Aboyoun P, Pages H, et al. ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. Bioinformatics 2009; 25: 2607–2608. doi: 10.1093/bioinformatics/btp450 PMID: 19654119

90. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 2009; 10: R25. doi: 10.1186/gb-2009-10-3-r25 PMID: 19261174

91. de Hoon MJ, Taft RJ, Hashimoto T, Kanamori-Katayama M, Kawaji H, et al. Cross-mapping and the identification of editing sites in mature microRNAs in high-throughput sequencing libraries. Genome Res 2010; 20: 257–264. doi: 10.1101/gr.095273.109 PMID: 20051556

92. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010; 26: 841–842. doi: 10.1093/bioinformatics/btq033 PMID: 20110278

93. Fournier D, Skaug H, Ancheta J, Ianelli J, Magnusson A, et al. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optimization Methods and Software 2012; 27: 233–249.

94. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. Nature 2012; 491: 56–65. doi: 10.1038/nature11632 PMID: 23128226

95. Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 1992; 131: 479–491. PMID: 1644282

96. Weir CL, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution 1984; 38: 1358–1370.

97. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, et al. Global variation in copy number in the human genome. Nature 2006; 444: 444–454. PMID: 17122850

98. Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, et al. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell 2007; 27: 91–105. PMID: 17612493

99. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 2008; 9: 559. doi: 10.1186/1471-2105-9-559 PMID: 19114008

100. Langfelder P, Horvath S. Fast R Functions for Robust Correlations and Hierarchical Clustering. J Stat Softw 2012; 46. PMID: 23050260

101. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 2001; 25: 402–408. PMID: 11846609

# 2. Complément d'informations pour l'article 1

# Genomic Signatures of Selective Pressures and Introgression from Archaic Hominins at Human Innate Immunity Genes

Matthieu Deschamps,[1,2,3] Guillaume Laval,[1,2] Maud Fagny,[1,2,3], Yuval Itan,[4] Laurent Abel,[4,5,6] Jean-Laurent Casanova,[4,5,6] Etienne Patin,[1,2] and Lluis Quintana-Murci[1,2,*]

[1]Institut Pasteur, Unit of Human Evolutionary Genetics, 75015 Paris, France

[2]CNRS URA3012, 75015 Paris, France

[3]Université Pierre et Marie Curie, Cellule Pasteur UPMC, 75015 Paris, France

[4]St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, The Rockefeller University, New York, NY 10065, USA

[5]Laboratory of Human Genetics of Infectious Diseases, Necker Branch, INSERM U.1163, 7515 Paris, France

[6]University Paris Descartes, Imagine Institute, 75015 Paris, France

*Correspondence: quintana@pasteur.fr

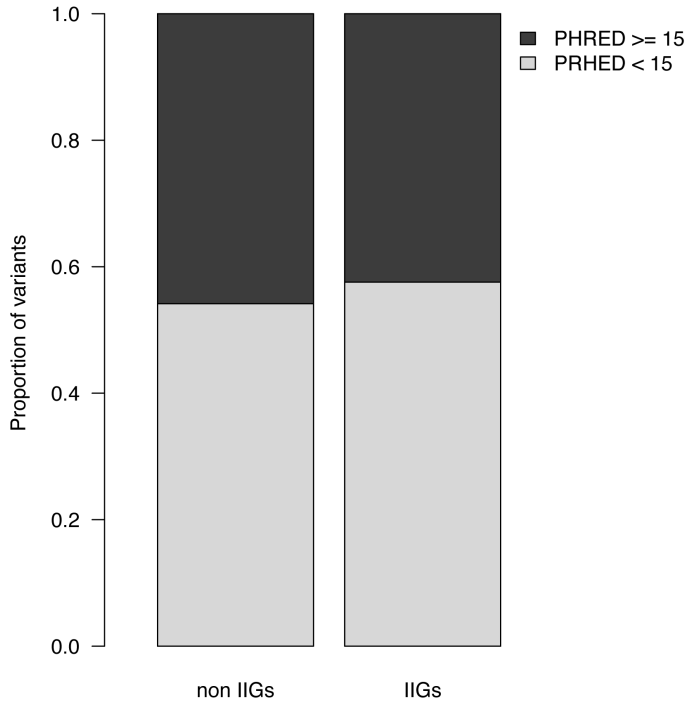



**Figure S1 Proportion of damaging and benign mutations in non-IIGs and IIGs as assessed by PHRED-scaled C-scores calculated using the CADD algorithm.** We considered a score of 15 as the limit above which mutations are probably damaging, as this value corresponds to the median value for all possible canonical splice site changes and non-synonymous variants.

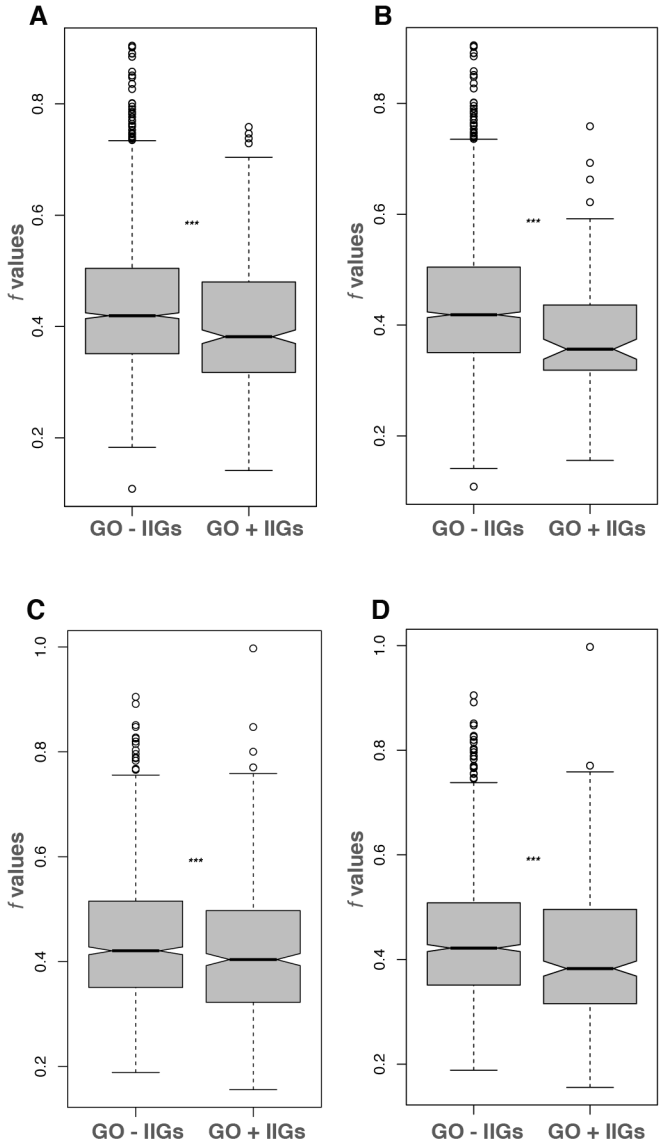



**Figure S2 Intensity of purifying selection acting on genes involved in transcription and signal transduction processes.** (**A**) Distribution of *f* values obtained for our IIG list that are also reported in the Gene Ontology term "transcription, DNA-templated". (**B**) For the analysis of transcription, to test if this observation was influenced by the genes we added as well as those present only in InnateDB, we performed the same comparison using Gene Ontology "innate immune response" gene list. (**C**) Distribution of *f* values obtained for our IIG list that are also reported in the Gene Ontology term "intracellular signal transduction". (**D**) For the analysis of signal transduction, to test if this observation was influenced by the genes we added as well as those present only in InnateDB, we performed the same comparison using Gene Ontology "innate immune response" gene list. ***: Wilcoxon $P<0.001$

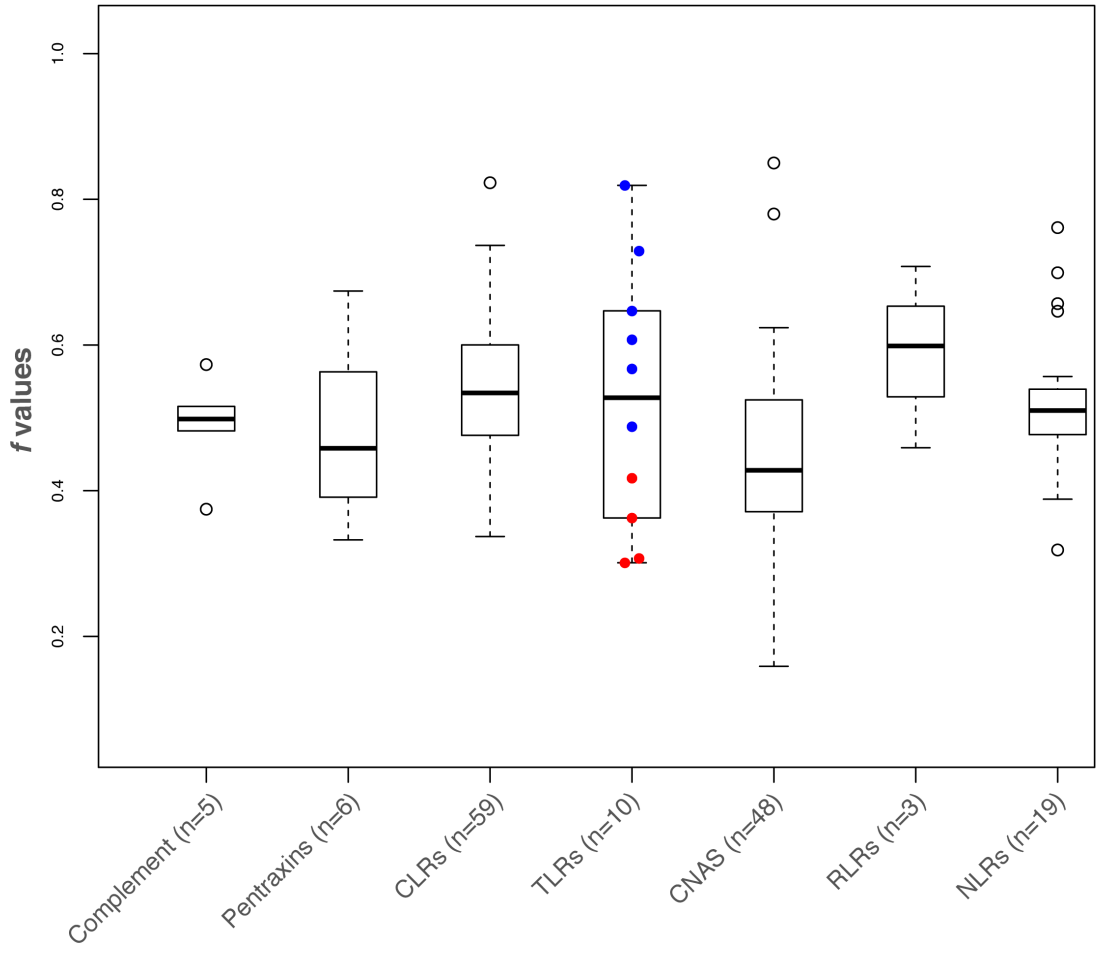



**Figure S3 Estimation of the intensity of purifying selection acting on major families of innate immunity receptors.** For Toll-Like Receptors, red dots correspond to the *f* values obtained for endosomal TLRs, which sense nucleic acids, whereas blue dots correspond to the *f* values obtained for TLRs expressed at the cell surface. CLRs: C-type Lectin Receptors; TLRs: Toll-Like Receptors; CNAS: Cytosolic Nucleic Acid Sensors; RLRs: RIG-I Like Receptors; NLRs: NOD-Like Receptors

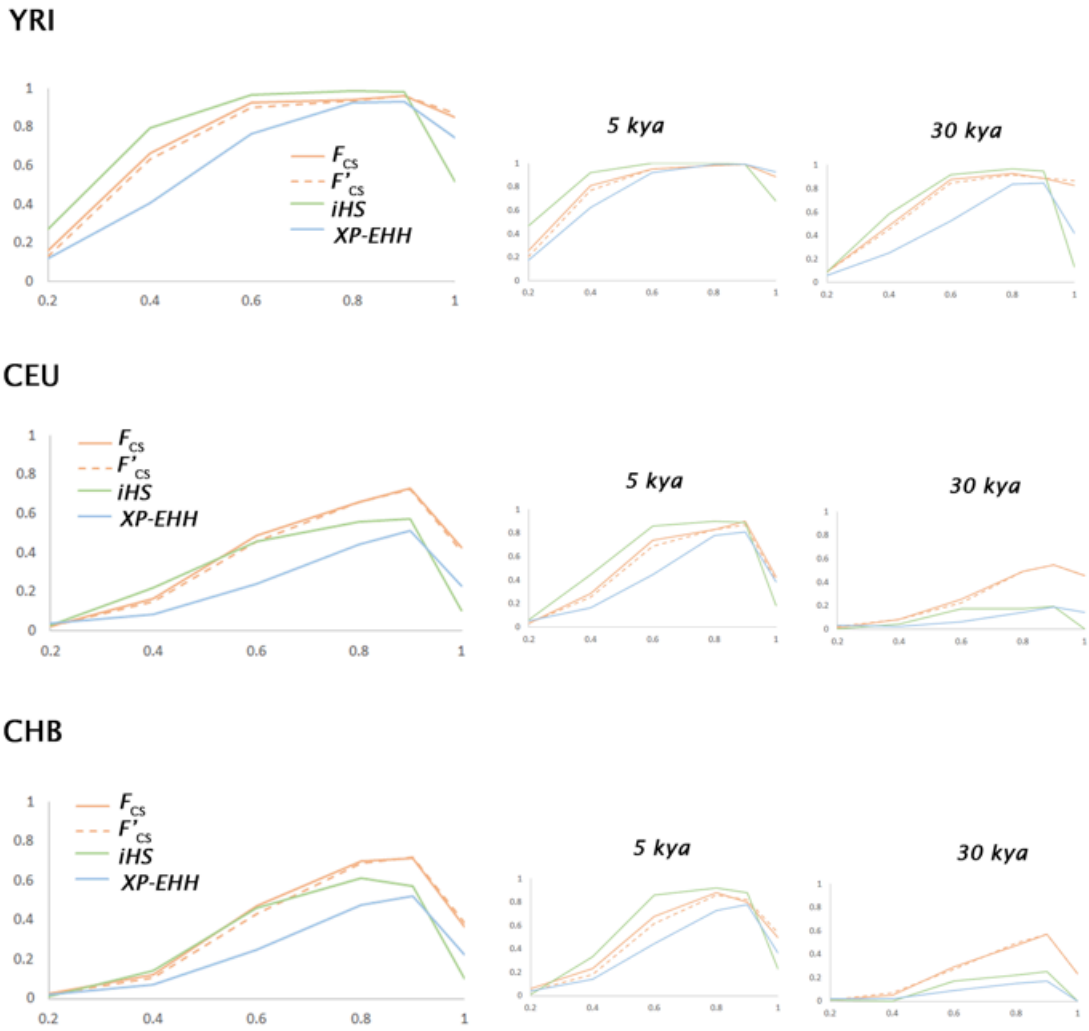



**Figure S4 Power of the $F_{CS}$ combined statistics to detect positive selection.** We simulated 200 kb DNA regions according to the realistic demography of West African (YRI), European (CEU) and East Asian (CHB) samples (Material and Methods). We simulated positive selection models in each population separately, using various age $t$ of the selected allele (5 kya, 10 kya, 20 kya and 30 kya), current frequency of the selected allele $p_{sel}$ (0.2, 0.4, 0.6, 0.8, and 1.0) and setting the selection coefficient $s$ to be equal to 0.01 (100 datasets for each parameter combination, see Material and Methods). For each population, we plot the power (proportion of simulated regions under selection effectively detected) obtained using the $F_{CS}$ and $F'_{CS}$ combined statistics, i. e., with or without DIND statistics, respectively (Material and Methods, FPR of 1%). We indicate the power of classical statistics, iHS and XP-EHH. Left panel shows the power obtained with age of selection $t$ uniformly distributed from 5 kya to 30 kya. Middle and right panels show the power obtained when the selected allele occurred 5 kya and 30 kya, respectively. The $x$-axis represents the current frequency of the selected allele $p_{sel}$.

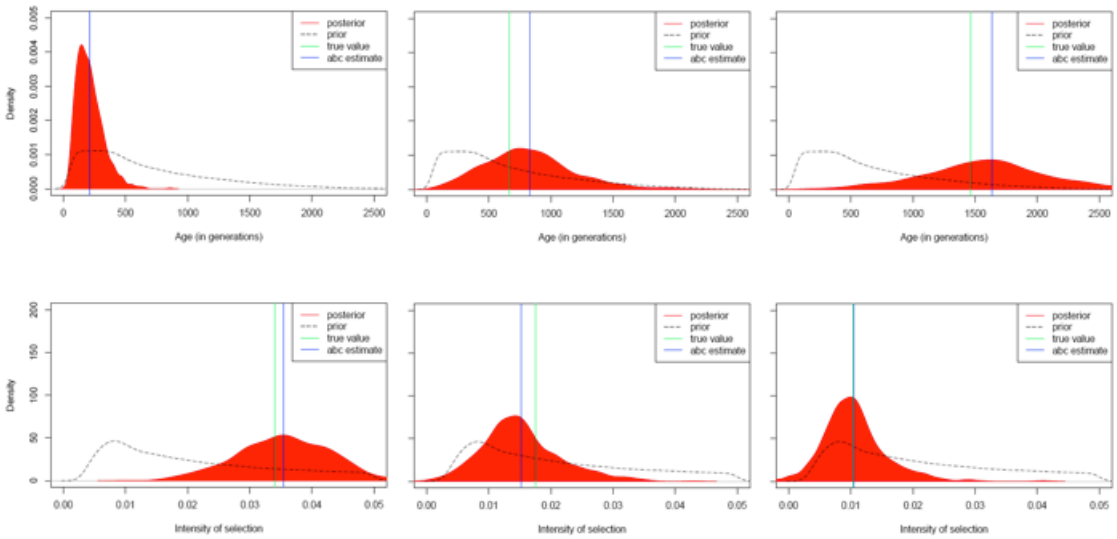



**Figure S5 Posterior distribution of the age and intensity of selection obtained using simulated data.** This figure shows some examples of posterior distributions estimated using the 'boosted-neuralnet' ABC method (Table S5). Black dotted lines are the prior distributions. Red surfaces are the posterior distributions. The true values are indicated in green and the punctual ABC estimates (posterior mean) are given in blue (true values, punctual estimates and 95% credible intervals are also given in Table S6).

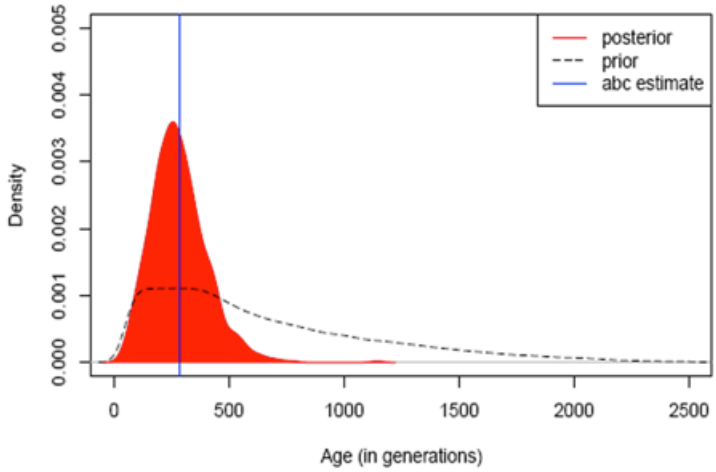



**Figure S6 Posterior distribution of the age of selection estimated for the variant rs4988235 in the *LCT* region.** This figure shows the posterior distribution corresponding to the 'boosted-neuralnet' ABC estimations given in Table S5. Black dotted line is the prior distribution. Red surface is the posterior distribution. The ABC estimate (posterior mean) is given in blue (the punctual estimation and the 95% credible interval are given in Table S7).

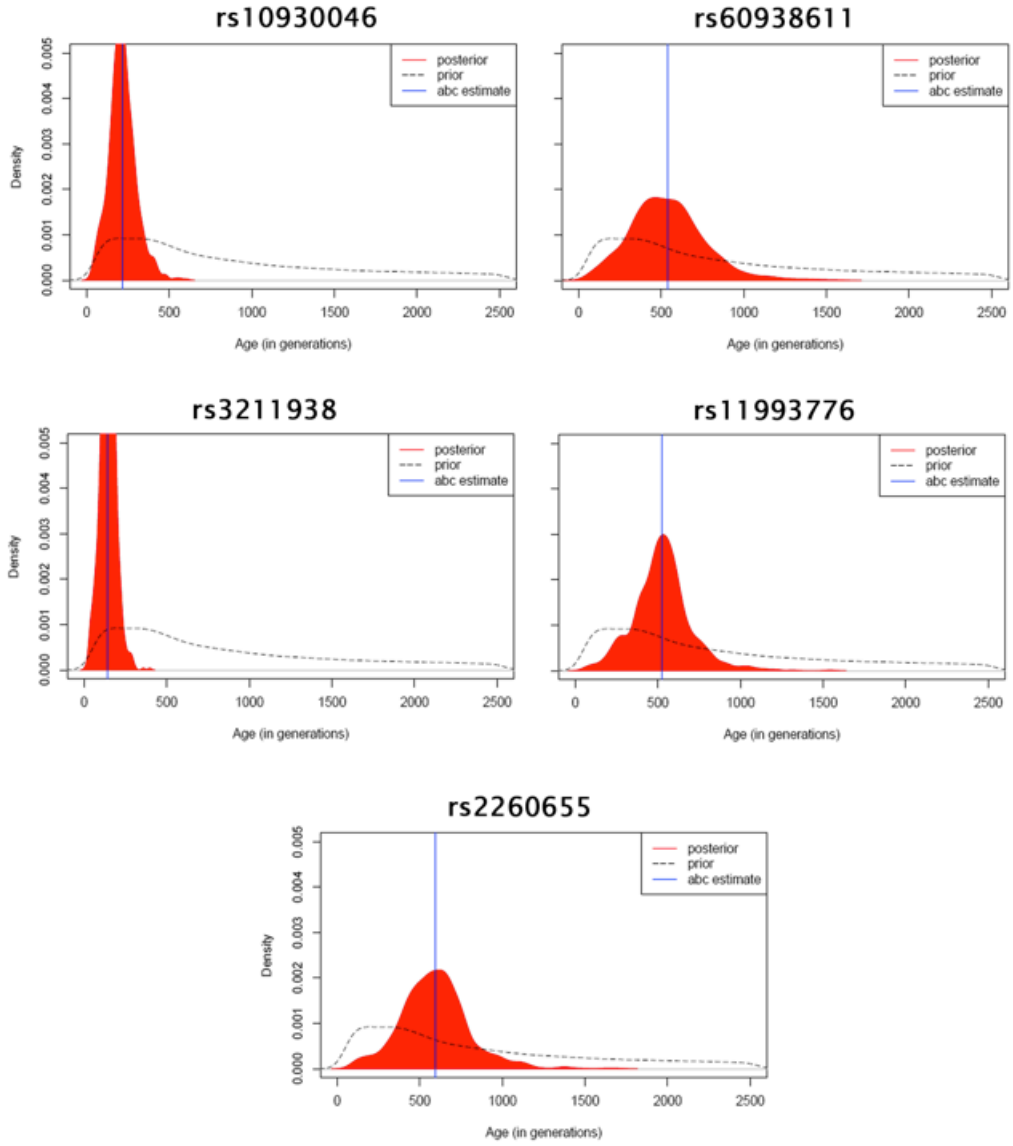



**Figure S7 Posterior distribution of the ages of selection in the YRI sample.** This figure shows the posterior distributions corresponding to the 'boosted-neuralnet' ABC estimations given in Table S5. Black dotted lines are the prior distributions. Red surfaces are the posterior distributions. The ABC estimates (posterior mean) are given in blue (punctual estimations and 95% credible intervals are given in Table S7).

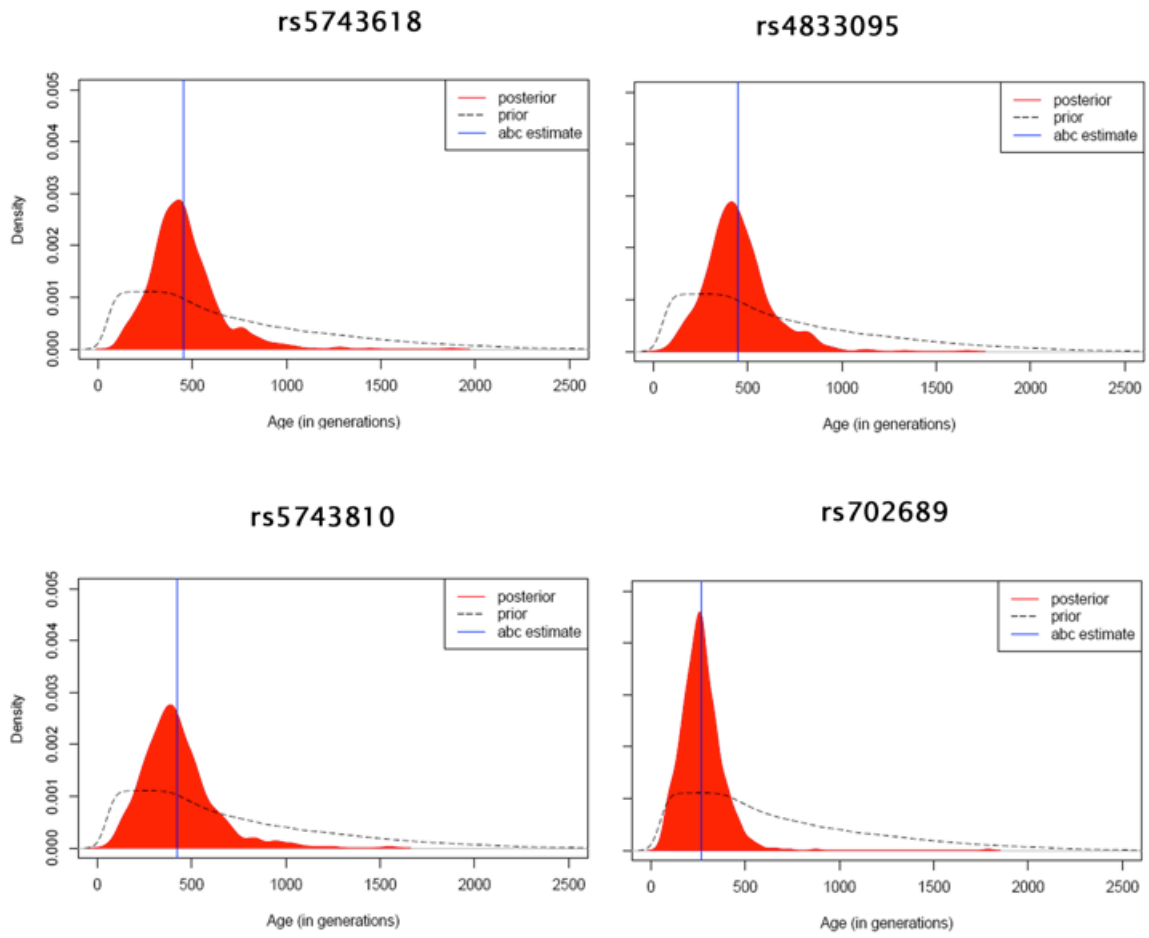



**Figure S8 Posterior distributions of the ages of selection in the CEU sample.** This figure shows the posterior distributions corresponding to the 'boosted-neuralnet' ABC estimations given in Table S5. Black dotted lines are the prior distributions. Red surfaces are the posterior distributions. The ABC estimates (posterior mean) are given in blue (punctual estimations and 95% credible intervals are given in Table S7).

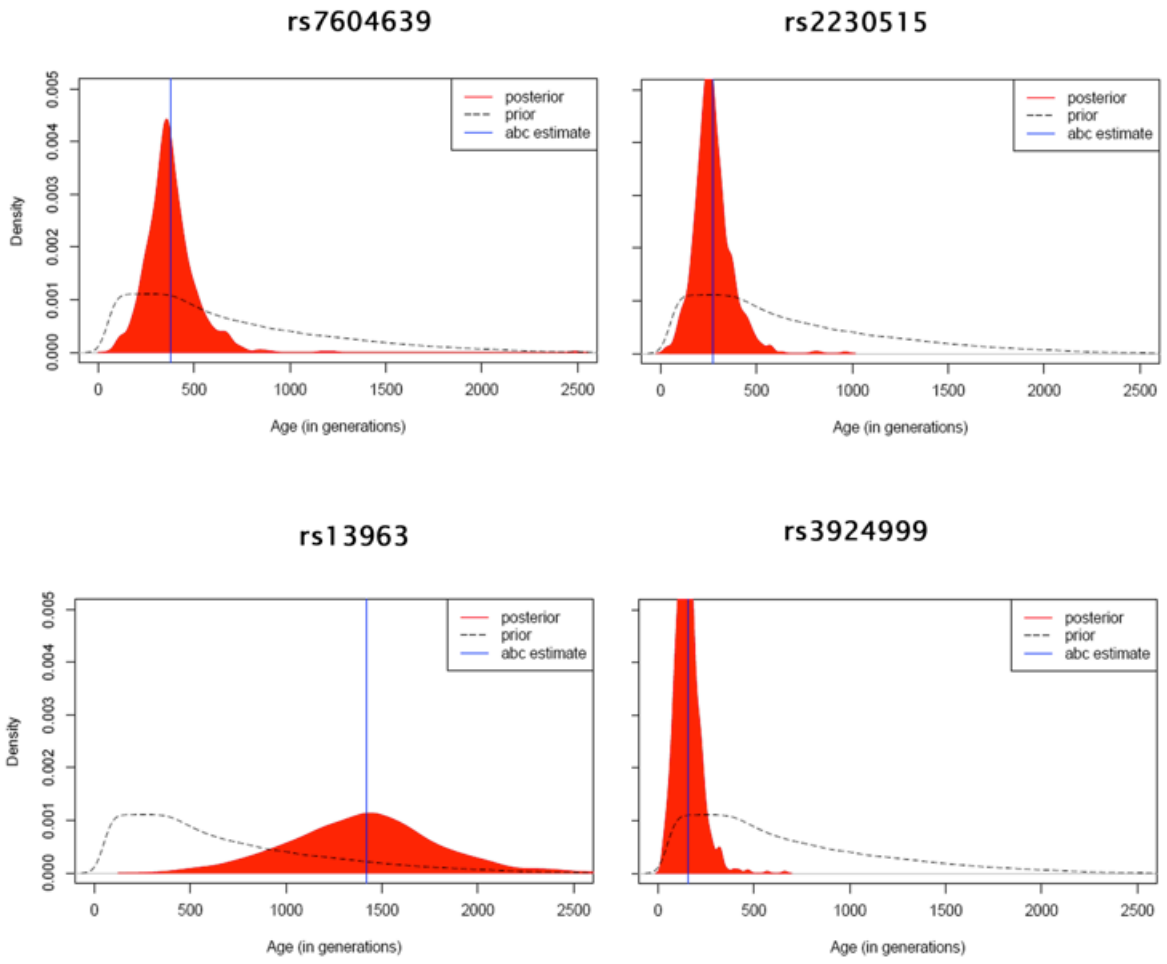



**Figure S9 Posterior distributions of the ages of selection in the CHB sample.** This figure shows the posterior distributions corresponding to the 'boosted-neuralnet' ABC estimations given in Table S5. Black dotted lines are the prior distributions. Red surfaces are the posterior distributions. The ABC estimates (posterior mean) are given in blue (punctual estimations and 95% credible intervals are given in Table S7).

**Table S1.** Full list of innate immunity genes

##HGNC_symbol: official gene Symbol

##Description: full protein name

##IIG_class: classification of innate immunity genes (levels: receptor, adaptor, signal transducer, transcription, effector, secondary receptor, regulator, accessory molecule, UC)

##GO: is it included in Gene Ontology "innate immunity response" gene list? (yes 1/0 no)

##InnateDB: is it included in InnateDB gene list? (yes 1/0 no)

##f_value: corresponding f value calculated using SnIPRE. NA means the gene was removed from analysis (see Material and Methods)

| HGNC_symbol | Description | IGG_class | GO | InnateDB | f_value |
|---|---|---|---|---|---|
| SMARCA4 | SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4 [Source:HGNC Symbol;Acc:HGNC:11100] | transcription | 0 | 1 | 0.141334209 |
| AGO2 | argonaute RISC catalytic component 2 [Source:HGNC Symbol;Acc:HGNC:3263] | effector | 1 | 0 | 0.155772933 |
| DHX9 | DEAH (Asp-Glu-Ala-His) box helicase 9 [Source:HGNC Symbol;Acc:HGNC:2750] | receptor | 1 | 1 | 0.159014684 |
| CYFIP2 | cytoplasmic FMR1 interacting protein 2 [Source:HGNC Symbol;Acc:HGNC:13760] | UC | 1 | 0 | 0.167738586 |
| ACTG1 | actin gamma 1 [Source:HGNC Symbol;Acc:HGNC:144] | UC | 1 | 0 | 0.174174069 |
| UBC | ubiquitin C [Source:HGNC Symbol;Acc:HGNC:12468] | accessory molecule | 1 | 0 | 0.174270539 |
| ITPR1 | inositol 1,4,5-trisphosphate receptor, type 1 [Source:HGNC Symbol;Acc:HGNC:6180] | UC | 1 | 1 | 0.176032366 |
| KHSRP | KH-type splicing regulatory protein [Source:HGNC Symbol;Acc:HGNC:6316] | regulator | 0 | 1 | 0.185959369 |
| TUBB4B | tubulin, beta 4B class IVb [Source:HGNC Symbol;Acc:HGNC:20771] | UC | 1 | 0 | 0.192025298 |
| STAT1 | signal transducer and activator of transcription 1, 91kDa [Source:HGNC Symbol;Acc:HGNC:11362] | transcription | 1 | 1 | 0.194396545 |
| AGO3 | argonaute RISC catalytic component 3 [Source:HGNC Symbol;Acc:HGNC:18421] | effector | 1 | 0 | 0.200181534 |
| ACTB | actin, beta [Source:HGNC Symbol;Acc:HGNC:132] | UC | 1 | 0 | 0.204873307 |
| HNRNPL | heterogeneous nuclear ribonucleoprotein L [Source:HGNC Symbol;Acc:HGNC:5045] | transcription | 0 | 1 | 0.208137015 |
| USP7 | ubiquitin specific peptidase 7 (herpes virus-associated) [Source:HGNC Symbol;Acc:HGNC:12630] | regulator | 0 | 1 | 0.209014199 |
| GATA3 | GATA binding protein 3 [Source:HGNC Symbol;Acc:HGNC:4172] | transcription | 1 | 0 | 0.210563348 |
| SRC | SRC proto-oncogene, non-receptor tyrosine kinase [Source:HGNC Symbol;Acc:HGNC:11283] | adaptor | 1 | 1 | 0.211953612 |
| MTOR | mechanistic target of rapamycin (serine/threonine kinase) [Source:HGNC Symbol;Acc:HGNC:3942] | signal transducer | 1 | 1 | 0.211957039 |
| CYLD | cylindromatosis (turban tumor syndrome) [Source:HGNC Symbol;Acc:HGNC:2584] | regulator | 1 | 1 | 0.214175105 |

| | | | | | |
|---|---|---|---|---|---|
| SMARCA2 | SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2 [Source:HGNC Symbol;Acc:HGNC:11098] | transcription | 0 | 1 | 0.215929494 |
| FSCN1 | fascin actin-bundling protein 1 [Source:HGNC Symbol;Acc:HGNC:11148] | UC | 0 | 1 | 0.216847421 |
| DOCK1 | dedicator of cytokinesis 1 [Source:HGNC Symbol;Acc:HGNC:2987] | UC | 1 | 0 | 0.218371744 |
| NCKAP1 | NCK-associated protein 1 [Source:HGNC Symbol;Acc:HGNC:7666] | UC | 1 | 0 | 0.219499405 |
| EGFR | epidermal growth factor receptor [Source:HGNC Symbol;Acc:HGNC:3236] | secondary receptor | 1 | 1 | 0.221079694 |
| CTNNB1 | catenin (cadherin-associated protein), beta 1, 88kDa [Source:HGNC Symbol;Acc:HGNC:2514] | adaptor | 1 | 1 | 0.221098897 |
| CAMK2B | calcium/calmodulin-dependent protein kinase II beta [Source:HGNC Symbol;Acc:HGNC:1461] | signal transducer | 1 | 0 | 0.224617801 |
| SYK | spleen tyrosine kinase [Source:HGNC Symbol;Acc:HGNC:11491] | adaptor | 1 | 0 | 0.225958822 |
| MID2 | midline 2 [Source:HGNC Symbol;Acc:HGNC:7096] | regulator | 1 | 1 | 0.227663853 |
| AGO1 | argonaute RISC catalytic component 1 [Source:HGNC Symbol;Acc:HGNC:3262] | effector | 1 | 0 | 0.228134346 |
| HDAC1 | histone deacetylase 1 [Source:HGNC Symbol;Acc:HGNC:4852] | regulator | 0 | 1 | 0.228712929 |
| CNKSR2 | connector enhancer of kinase suppressor of Ras 2 [Source:HGNC Symbol;Acc:HGNC:19701] | adaptor | 0 | 0 | 0.228828954 |
| HSP90AB1 | heat shock protein 90kDa alpha (cytosolic), class B member 1 [Source:HGNC Symbol;Acc:HGNC:5258] | accessory molecule | 1 | 0 | 0.231202738 |
| TRAF3 | TNF receptor-associated factor 3 [Source:HGNC Symbol;Acc:HGNC:12033] | signal transducer | 1 | 1 | 0.233702447 |
| LYN | LYN proto-oncogene, Src family tyrosine kinase [Source:HGNC Symbol;Acc:HGNC:6735] | signal transducer | 1 | 1 | 0.23712539 |
| ANKRD17 | ankyrin repeat domain 17 [Source:HGNC Symbol;Acc:HGNC:23575] | UC | 1 | 1 | 0.237926748 |
| MAP2K7 | mitogen-activated protein kinase kinase 7 [Source:HGNC Symbol;Acc:HGNC:6847] | signal transducer | 1 | 1 | 0.239022429 |
| DUSP4 | dual specificity phosphatase 4 [Source:HGNC Symbol;Acc:HGNC:3070] | regulator | 1 | 0 | 0.240032438 |
| ADAM10 | ADAM metallopeptidase domain 10 [Source:HGNC Symbol;Acc:HGNC:188] | regulator | 0 | 1 | 0.240981107 |
| KDM1A | lysine (K)-specific demethylase 1A [Source:HGNC Symbol;Acc:HGNC:29079] | regulator | 0 | 0 | 0.241540502 |
| DHX15 | DEAH (Asp-Glu-Ala-His) box helicase 15 [Source:HGNC Symbol;Acc:HGNC:2738] | receptor | 0 | 0 | 0.242341787 |
| ZMYND11 | zinc finger, MYND-type containing 11 [Source:HGNC Symbol;Acc:HGNC:16966] | regulator | 0 | 1 | 0.243032742 |
| RNF19B | ring finger protein 19B [Source:HGNC Symbol;Acc:HGNC:26886] | effector | 1 | 0 | 0.243329438 |
| BAIAP2 | BAI1-associated protein 2 [Source:HGNC Symbol;Acc:HGNC:947] | adaptor | 1 | 0 | 0.243653056 |
| PRKAR1A | protein kinase, cAMP-dependent, regulatory, type I, alpha [Source:HGNC Symbol;Acc:HGNC:9388] | regulator | 1 | 0 | 0.243894533 |
| PPP2R5D | protein phosphatase 2, regulatory subunit B', delta [Source:HGNC Symbol;Acc:HGNC:9312] | regulator | 1 | 0 | 0.244746839 |

| | | | | | |
|---|---|---|---|---|---|
| TRIM67 | tripartite motif containing 67 [Source:HGNC Symbol;Acc:HGNC:31859] | regulator | 0 | 1 | 0.245039289 |
| MAPK10 | mitogen-activated protein kinase 10 [Source:HGNC Symbol;Acc:HGNC:6872] | signal transducer | 1 | 0 | 0.24603225 |
| HSP90AA1 | heat shock protein 90kDa alpha (cytosolic), class A member 1 [Source:HGNC Symbol;Acc:HGNC:5253] | accessory molecule | 1 | 1 | 0.24741086 |
| CAMK2A | calcium/calmodulin-dependent protein kinase II alpha [Source:HGNC Symbol;Acc:HGNC:1460] | signal transducer | 1 | 1 | 0.248762286 |
| ELAVL1 | ELAV like RNA binding protein 1 [Source:HGNC Symbol;Acc:HGNC:3312] | UC | 0 | 1 | 0.251300386 |
| MTA1 | metastasis associated 1 [Source:HGNC Symbol;Acc:HGNC:7410] | transcription | 0 | 1 | 0.251403434 |
| CLTC | clathrin, heavy chain (Hc) [Source:HGNC Symbol;Acc:HGNC:2092] | regulator | 0 | 1 | 0.251480805 |
| ZAP70 | zeta-chain (TCR) associated protein kinase 70kDa [Source:HGNC Symbol;Acc:HGNC:12858] | signal transducer | 1 | 0 | 0.252148592 |
| IL1RAP | interleukin 1 receptor accessory protein [Source:HGNC Symbol;Acc:HGNC:5995] | secondary receptor | 1 | 1 | 0.25272467 |
| TRIM2 | tripartite motif containing 2 [Source:HGNC Symbol;Acc:HGNC:15974] | regulator | 0 | 0 | 0.25305206 |
| CDK9 | cyclin-dependent kinase 9 [Source:HGNC Symbol;Acc:HGNC:1780] | regulator | 0 | 1 | 0.254739795 |
| EHMT2 | euchromatic histone-lysine N-methyltransferase 2 [Source:HGNC Symbol;Acc:HGNC:14129] | UC | 0 | 1 | 0.255633446 |
| TAB2 | TGF-beta activated kinase 1/MAP3K7 binding protein 2 [Source:HGNC Symbol;Acc:HGNC:17075] | signal transducer | 1 | 0 | 0.255784496 |
| SOCS5 | suppressor of cytokine signaling 5 [Source:HGNC Symbol;Acc:HGNC:16852] | effector | 0 | 1 | 0.255917769 |
| CAMK2G | calcium/calmodulin-dependent protein kinase II gamma [Source:HGNC Symbol;Acc:HGNC:1463] | signal transducer | 1 | 0 | 0.257070633 |
| GNB2 | guanine nucleotide binding protein (G protein), beta polypeptide 2 [Source:HGNC Symbol;Acc:HGNC:4398] | UC | 0 | 1 | 0.257269172 |
| TUBB | tubulin, beta class I [Source:HGNC Symbol;Acc:HGNC:20778] | UC | 1 | 0 | 0.257339256 |
| PPP1CA | protein phosphatase 1, catalytic subunit, alpha isozyme [Source:HGNC Symbol;Acc:HGNC:9281] | accessory molecule | 0 | 1 | 0.257995002 |
| IPO7 | importin 7 [Source:HGNC Symbol;Acc:HGNC:9852] | accessory molecule | 1 | 0 | 0.258386272 |
| MAPK1 | mitogen-activated protein kinase 1 [Source:HGNC Symbol;Acc:HGNC:6871] | signal transducer | 1 | 1 | 0.2586921 |
| STAT5A | signal transducer and activator of transcription 5A [Source:HGNC Symbol;Acc:HGNC:11366] | transcription | 0 | 1 | 0.260118374 |
| PTEN | phosphatase and tensin homolog [Source:HGNC Symbol;Acc:HGNC:9588] | regulator | 1 | 1 | 0.261082584 |
| JUN | jun proto-oncogene [Source:HGNC Symbol;Acc:HGNC:6204] | transcription | 1 | 1 | 0.261446287 |

| | | | | | |
|---|---|---|---|---|---|
| NXN | nucleoredoxin [Source:HGNC Symbol;Acc:HGNC:18008] | signal transducer | 0 | 1 | 0.262017178 |
| SRPK2 | SRSF protein kinase 2 [Source:HGNC Symbol;Acc:HGNC:11306] | signal transducer | 1 | 0 | 0.262357478 |
| OTUD5 | OTU deubiquitinase 5 [Source:HGNC Symbol;Acc:HGNC:25402] | regulator | 1 | 1 | 0.262784544 |
| STAT3 | signal transducer and activator of transcription 3 (acute-phase response factor) [Source:HGNC Symbol;Acc:HGNC:11364] | transcription | 0 | 1 | 0.263077493 |
| TRIM71 | tripartite motif containing 71, E3 ubiquitin protein ligase [Source:HGNC Symbol;Acc:HGNC:32669] | regulator | 0 | 1 | 0.263997399 |
| HSPA1B | heat shock 70kDa protein 1B [Source:HGNC Symbol;Acc:HGNC:5233] | effector | 0 | 0 | 0.264620831 |
| PIAS4 | protein inhibitor of activated STAT, 4 [Source:HGNC Symbol;Acc:HGNC:17002] | regulator | 0 | 1 | 0.26480869 |
| TAB3 | TGF-beta activated kinase 1/MAP3K7 binding protein 3 [Source:HGNC Symbol;Acc:HGNC:30681] | signal transducer | 1 | 0 | 0.265429923 |
| AKT1 | v-akt murine thymoma viral oncogene homolog 1 [Source:HGNC Symbol;Acc:HGNC:391] | signal transducer | 1 | 1 | 0.266376018 |
| PAK2 | p21 protein (Cdc42/Rac)-activated kinase 2 [Source:HGNC Symbol;Acc:HGNC:8591] | signal transducer | 1 | 0 | 0.266750713 |
| PRKCD | protein kinase C, delta [Source:HGNC Symbol;Acc:HGNC:9399] | signal transducer | 1 | 1 | 0.267271827 |
| DDX41 | DEAD (Asp-Glu-Ala-Asp) box polypeptide 41 [Source:HGNC Symbol;Acc:HGNC:18674] | receptor | 1 | 1 | 0.267409262 |
| PPP3CA | protein phosphatase 3, catalytic subunit, alpha isozyme [Source:HGNC Symbol;Acc:HGNC:9314] | signal transducer | 1 | 1 | 0.268457817 |
| INPP5D | inositol polyphosphate-5-phosphatase, 145kDa [Source:HGNC Symbol;Acc:HGNC:6079] | signal transducer | 0 | 1 | 0.269512011 |
| ADCY1 | adenylate cyclase 1 (brain) [Source:HGNC Symbol;Acc:HGNC:232] | UC | 1 | 0 | 0.270902847 |
| DAPK1 | death-associated protein kinase 1 [Source:HGNC Symbol;Acc:HGNC:2674] | signal transducer | 1 | 0 | 0.271465188 |
| BTK | Bruton agammaglobulinemia tyrosine kinase [Source:HGNC Symbol;Acc:HGNC:1133] | signal transducer | 1 | 1 | 0.271653018 |
| PIAS1 | protein inhibitor of activated STAT, 1 [Source:HGNC Symbol;Acc:HGNC:2752] | regulator | 1 | 1 | 0.272342249 |
| PCBP2 | poly(rC) binding protein 2 [Source:HGNC Symbol;Acc:HGNC:8648] | adaptor | 1 | 1 | 0.273551726 |
| TRIM33 | tripartite motif containing 33 [Source:HGNC Symbol;Acc:HGNC:16290] | regulator | 0 | 0 | 0.273562663 |
| RAC1 | ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1) [Source:HGNC Symbol;Acc:HGNC:9801] | regulator | 1 | 1 | 0.273758415 |
| HMGB2 | high mobility group box 2 [Source:HGNC Symbol;Acc:HGNC:5000] | receptor | 0 | 1 | 0.274299911 |

| | | | | | |
|---|---|---|---|---|---|
| PIK3C3 | phosphatidylinositol 3-kinase, catalytic subunit type 3 [Source:HGNC Symbol;Acc:HGNC:8974] | signal transducer | 1 | 1 | 0.275106618 |
| MOV10 | Mov10 RISC complex RNA helicase [Source:HGNC Symbol;Acc:HGNC:7200] | effector | 1 | 1 | 0.276795252 |
| PTPN6 | protein tyrosine phosphatase, non-receptor type 6 [Source:HGNC Symbol;Acc:HGNC:9658] | regulator | 1 | 1 | 0.277111401 |
| MEF2C | myocyte enhancer factor 2C [Source:HGNC Symbol;Acc:HGNC:6996] | transcription | 1 | 0 | 0.277479579 |
| YWHAE | tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, epsilon [Source:HGNC Symbol;Acc:HGNC:12851] | signal transducer | 0 | 1 | 0.277518681 |
| CSK | c-src tyrosine kinase [Source:HGNC Symbol;Acc:HGNC:2444] | signal transducer | 1 | 1 | 0.277888725 |
| CDK1 | cyclin-dependent kinase 1 [Source:HGNC Symbol;Acc:HGNC:1722] | signal transducer | 1 | 0 | 0.278511161 |
| KLF4 | Kruppel-like factor 4 (gut) [Source:HGNC Symbol;Acc:HGNC:6348] | transcription | 0 | 0 | 0.278872323 |
| CREBBP | CREB binding protein [Source:HGNC Symbol;Acc:HGNC:2348] | transcription | 1 | 0 | 0.27895951 |
| DDX3X | DEAD (Asp-Glu-Ala-Asp) box helicase 3, X-linked [Source:HGNC Symbol;Acc:HGNC:2745] | receptor | 1 | 1 | 0.279086644 |
| ELMO1 | engulfment and cell motility 1 [Source:HGNC Symbol;Acc:HGNC:16286] | UC | 1 | 0 | 0.279415779 |
| CCDC88A | coiled-coil domain containing 88A [Source:HGNC Symbol;Acc:HGNC:25523] | regulator | 0 | 1 | 0.279477136 |
| ETS1 | v-ets avian erythroblastosis virus E26 oncogene homolog 1 [Source:HGNC Symbol;Acc:HGNC:3488] | transcription | 0 | 1 | 0.279520316 |
| IGF1R | insulin-like growth factor 1 receptor [Source:HGNC Symbol;Acc:HGNC:5465] | secondary receptor | 0 | 1 | 0.280128534 |
| ITPR3 | inositol 1,4,5-trisphosphate receptor, type 3 [Source:HGNC Symbol;Acc:HGNC:6182] | UC | 1 | 1 | 0.280377071 |
| CREB1 | cAMP responsive element binding protein 1 [Source:HGNC Symbol;Acc:HGNC:2345] | transcription | 1 | 1 | 0.28097779 |
| STAT5B | signal transducer and activator of transcription 5B [Source:HGNC Symbol;Acc:HGNC:11367] | transcription | 0 | 1 | 0.281347858 |
| PIK3CD | phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit delta [Source:HGNC Symbol;Acc:HGNC:8977] | signal transducer | 1 | 0 | 0.281484436 |
| RPS6KA3 | ribosomal protein S6 kinase, 90kDa, polypeptide 3 [Source:HGNC Symbol;Acc:HGNC:10432] | signal transducer | 1 | 0 | 0.281702399 |
| MIB2 | mindbomb E3 ubiquitin protein ligase 2 [Source:HGNC Symbol;Acc:HGNC:30577] | regulator | 0 | 0 | 0.281892858 |
| MYH9 | myosin, heavy chain 9, non-muscle [Source:HGNC Symbol;Acc:HGNC:7579] | UC | 0 | 1 | 0.281917325 |
| RXRA | retinoid X receptor, alpha [Source:HGNC Symbol;Acc:HGNC:10477] | receptor | 0 | 1 | 0.282184206 |
| ORAI1 | ORAI calcium release-activated calcium modulator 1 [Source:HGNC Symbol;Acc:HGNC:25896] | UC | 0 | 1 | 0.28237902 |
| PPP4C | protein phosphatase 4, catalytic subunit [Source:HGNC Symbol;Acc:HGNC:9319] | regulator | 0 | 1 | 0.282654805 |
| PAK1 | p21 protein (Cdc42/Rac)-activated kinase 1 [Source:HGNC Symbol;Acc:HGNC:8590] | signal transducer | 1 | 0 | 0.282806181 |

| | | | | | |
|---|---|---|---|---|---|
| AKT3 | v-akt murine thymoma viral oncogene homolog 3 [Source:HGNC Symbol;Acc:HGNC:393] | signal transducer | 0 | 1 | 0.282871689 |
| CTCF | CCCTC-binding factor (zinc finger protein) [Source:HGNC Symbol;Acc:HGNC:13723] | regulator | 0 | 1 | 0.283117071 |
| MAP2K1 | mitogen-activated protein kinase kinase 1 [Source:HGNC Symbol;Acc:HGNC:6840] | signal transducer | 1 | 1 | 0.283408831 |
| IKBKG | inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase gamma [Source:HGNC Symbol;Acc:HGNC:5961] | signal transducer | 1 | 1 | 0.28345019 |
| CEBPB | CCAAT/enhancer binding protein (C/EBP), beta [Source:HGNC Symbol;Acc:HGNC:1834] | transcription | 0 | 1 | 0.283630995 |
| XRCC6 | X-ray repair complementing defective repair in Chinese hamster cells 6 [Source:HGNC Symbol;Acc:HGNC:4055] | receptor | 1 | 1 | 0.284186306 |
| JAK1 | Janus kinase 1 [Source:HGNC Symbol;Acc:HGNC:6190] | signal transducer | 1 | 1 | 0.284502099 |
| SMAD7 | SMAD family member 7 [Source:HGNC Symbol;Acc:HGNC:6773] | signal transducer | 0 | 1 | 0.28479945 |
| ARHGEF7 | Rho guanine nucleotide exchange factor (GEF) 7 [Source:HGNC Symbol;Acc:HGNC:15607] | regulator | 0 | 0 | 0.285418172 |
| NAMPT | nicotinamide phosphoribosyltransferase [Source:HGNC Symbol;Acc:HGNC:30092] | effector | 0 | 1 | 0.285419061 |
| ITGB1 | integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12) [Source:HGNC Symbol;Acc:HGNC:6153] | receptor | 0 | 1 | 0.285639919 |
| RANBP9 | RAN binding protein 9 [Source:HGNC Symbol;Acc:HGNC:13727] | accessory molecule | 0 | 1 | 0.285896286 |
| FGR | FGR proto-oncogene, Src family tyrosine kinase [Source:HGNC Symbol;Acc:HGNC:3697] | signal transducer | 1 | 0 | 0.285958717 |
| MAP2K2 | mitogen-activated protein kinase kinase 2 [Source:HGNC Symbol;Acc:HGNC:6842] | signal transducer | 1 | 1 | 0.286367792 |
| MAPK14 | mitogen-activated protein kinase 14 [Source:HGNC Symbol;Acc:HGNC:6876] | signal transducer | 1 | 1 | 0.286981303 |
| EIF4E | eukaryotic translation initiation factor 4E [Source:HGNC Symbol;Acc:HGNC:3287] | UC | 0 | 1 | 0.287154716 |
| SMARCE1 | SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily e, member 1 [Source:HGNC Symbol;Acc:HGNC:11109] | transcription | 0 | 1 | 0.2873273 |
| RAD21 | RAD21 homolog (S. pombe) [Source:HGNC Symbol;Acc:HGNC:9811] | UC | 0 | 1 | 0.28733165 |
| PPP1CC | protein phosphatase 1, catalytic subunit, gamma isozyme [Source:HGNC Symbol;Acc:HGNC:9283] | accessory molecule | 0 | 1 | 0.287443895 |
| THRB | thyroid hormone receptor, beta [Source:HGNC Symbol;Acc:HGNC:11799] | effector | 0 | 1 | 0.287615384 |
| MLST8 | MTOR associated protein, LST8 homolog (S. cerevisiae) [Source:HGNC Symbol;Acc:HGNC:24825] | signal transducer | 1 | 1 | 0.288649529 |

| | | | | | |
|---|---|---|---|---|---|
| ATF2 | activating transcription factor 2 [Source:HGNC Symbol;Acc:HGNC:784] | transcription | 1 | 0 | 0.289144497 |
| PIK3CA | phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha [Source:HGNC Symbol;Acc:HGNC:8975] | signal transducer | 1 | 1 | 0.289469497 |
| ELMO2 | engulfment and cell motility 2 [Source:HGNC Symbol;Acc:HGNC:17233] | UC | 1 | 0 | 0.289585661 |
| SIAH2 | siah E3 ubiquitin protein ligase 2 [Source:HGNC Symbol;Acc:HGNC:10858] | signal transducer | 0 | 1 | 0.290051733 |
| SEC61A1 | Sec61 alpha 1 subunit (S. cerevisiae) [Source:HGNC Symbol;Acc:HGNC:18276] | accessory molecule | 1 | 0 | 0.290128337 |
| NEU1 | sialidase 1 (lysosomal sialidase) [Source:HGNC Symbol;Acc:HGNC:7758] | accessory molecule | 0 | 1 | 0.291217798 |
| IRF1 | interferon regulatory factor 1 [Source:HGNC Symbol;Acc:HGNC:6116] | transcription | 1 | 1 | 0.292167846 |
| YY1 | YY1 transcription factor [Source:HGNC Symbol;Acc:HGNC:12856] | regulator | 0 | 1 | 0.293676926 |
| PRKX | protein kinase, X-linked [Source:HGNC Symbol;Acc:HGNC:9441] | accessory molecule | 0 | 1 | 0.293853487 |
| RPS6KA2 | ribosomal protein S6 kinase, 90kDa, polypeptide 2 [Source:HGNC Symbol;Acc:HGNC:10431] | signal transducer | 1 | 0 | 0.294297531 |
| PRKAR1B | protein kinase, cAMP-dependent, regulatory, type I, beta [Source:HGNC Symbol;Acc:HGNC:9390] | regulator | 1 | 0 | 0.294507049 |
| KIT | v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog [Source:HGNC Symbol;Acc:HGNC:6342] | secondary receptor | 1 | 0 | 0.295129049 |
| NOTCH1 | notch 1 [Source:HGNC Symbol;Acc:HGNC:7881] | receptor | 0 | 1 | 0.295370451 |
| MID1 | midline 1 [Source:HGNC Symbol;Acc:HGNC:7095] | regulator | 0 | 1 | 0.295758465 |
| UBB | ubiquitin B [Source:HGNC Symbol;Acc:HGNC:12463] | accessory molecule | 1 | 0 | 0.295913258 |
| DTX4 | deltex 4, E3 ubiquitin ligase [Source:HGNC Symbol;Acc:HGNC:29151] | signal transducer | 1 | 0 | 0.296413087 |
| NR4A3 | nuclear receptor subfamily 4, group A, member 3 [Source:HGNC Symbol;Acc:HGNC:7982] | transcription | 0 | 1 | 0.296833045 |
| GRB2 | growth factor receptor-bound protein 2 [Source:HGNC Symbol;Acc:HGNC:4566] | secondary receptor | 1 | 0 | 0.297190176 |
| ADCY3 | adenylate cyclase 3 [Source:HGNC Symbol;Acc:HGNC:234] | UC | 1 | 0 | 0.298242957 |
| PRKCG | protein kinase C, gamma [Source:HGNC Symbol;Acc:HGNC:9402] | signal transducer | 1 | 0 | 0.298303114 |
| OTUB1 | OTU deubiquitinase, ubiquitin aldehyde binding 1 [Source:HGNC Symbol;Acc:HGNC:23077] | regulator | 0 | 1 | 0.298596538 |
| HSPD1 | heat shock 60kDa protein 1 (chaperonin) [Source:HGNC Symbol;Acc:HGNC:5261] | regulator | 0 | 1 | 0.299052116 |
| STAT4 | signal transducer and activator of transcription 4 [Source:HGNC Symbol;Acc:HGNC:11365] | transcription | 0 | 1 | 0.299754304 |

| | | | | | |
|---|---|---|---|---|---|
| TRIM26 | tripartite motif containing 26 [Source:HGNC Symbol;Acc:HGNC:12962] | regulator | 1 | 1 | 0.29991206 |
| TRIM27 | tripartite motif containing 27 [Source:HGNC Symbol;Acc:HGNC:9975] | regulator | 1 | 1 | 0.299970226 |
| GAPDH | glyceraldehyde-3-phosphate dehydrogenase [Source:HGNC Symbol;Acc:HGNC:4141] | signal transducer | 1 | 0 | 0.300614122 |
| TLR8 | toll-like receptor 8 [Source:HGNC Symbol;Acc:HGNC:15632] | receptor | 1 | 1 | 0.301103792 |
| NOS1 | nitric oxide synthase 1 (neuronal) [Source:HGNC Symbol;Acc:HGNC:7872] | effector | 0 | 0 | 0.301768014 |
| TRIM9 | tripartite motif containing 9 [Source:HGNC Symbol;Acc:HGNC:16288] | regulator | 0 | 1 | 0.302024231 |
| ARRB2 | arrestin, beta 2 [Source:HGNC Symbol;Acc:HGNC:712] | regulator | 0 | 1 | 0.302225281 |
| TEC | tec protein tyrosine kinase [Source:HGNC Symbol;Acc:HGNC:11719] | signal transducer | 1 | 0 | 0.303049976 |
| ITPR2 | inositol 1,4,5-trisphosphate receptor, type 2 [Source:HGNC Symbol;Acc:HGNC:6181] | UC | 1 | 0 | 0.303119262 |
| ACTR3 | ARP3 actin-related protein 3 homolog (yeast) [Source:HGNC Symbol;Acc:HGNC:170] | UC | 1 | 0 | 0.303254017 |
| CXCR4 | chemokine (C-X-C motif) receptor 4 [Source:HGNC Symbol;Acc:HGNC:2561] | secondary receptor | 0 | 1 | 0.303371644 |
| BMX | BMX non-receptor tyrosine kinase [Source:HGNC Symbol;Acc:HGNC:1079] | regulator | 1 | 1 | 0.303775377 |
| ADRBK1 | adrenergic, beta, receptor kinase 1 [Source:HGNC Symbol;Acc:HGNC:289] | regulator | 1 | 1 | 0.303871591 |
| GSK3B | glycogen synthase kinase 3 beta [Source:HGNC Symbol;Acc:HGNC:4617] | signal transducer | 1 | 1 | 0.304599937 |
| RPS6KB1 | ribosomal protein S6 kinase, 70kDa, polypeptide 1 [Source:HGNC Symbol;Acc:HGNC:10436] | signal transducer | 0 | 0 | 0.304665014 |
| PPP3CB | protein phosphatase 3, catalytic subunit, beta isozyme [Source:HGNC Symbol;Acc:HGNC:9315] | signal transducer | 1 | 0 | 0.304745731 |
| PELI1 | pellino E3 ubiquitin protein ligase 1 [Source:HGNC Symbol;Acc:HGNC:8827] | signal transducer | 1 | 1 | 0.305203468 |
| NTN1 | netrin 1 [Source:HGNC Symbol;Acc:HGNC:8029] | UC | 0 | 1 | 0.305754482 |
| AKIRIN2 | akirin 2 [Source:HGNC Symbol;Acc:HGNC:21407] | transcription | 1 | 0 | 0.305809657 |
| PLCG1 | phospholipase C, gamma 1 [Source:HGNC Symbol;Acc:HGNC:9065] | signal transducer | 1 | 0 | 0.305976029 |
| IKBKB | inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase beta [Source:HGNC Symbol;Acc:HGNC:5960] | signal transducer | 1 | 1 | 0.306207961 |
| NCKAP1L | NCK-associated protein 1-like [Source:HGNC Symbol;Acc:HGNC:4862] | UC | 0 | 1 | 0.306341229 |
| C4B | complement component 4B (Chido blood group) [Source:HGNC Symbol;Acc:HGNC:1324] | signal transducer | 1 | 1 | 0.306814986 |
| SYNCRIP | synaptotagmin binding, cytoplasmic RNA interacting protein [Source:HGNC | regulator | 1 | 0 | 0.306865282 |

| | | | | | |
|---|---|---|---|---|---|
| | Symbol;Acc:HGNC:16918] | | | | |
| TLR7 | toll-like receptor 7 [Source:HGNC Symbol;Acc:HGNC:15631] | receptor | 1 | 1 | 0.307184261 |
| PPIA | peptidylprolyl isomerase A (cyclophilin A) [Source:HGNC Symbol;Acc:HGNC:9253] | receptor | 0 | 1 | 0.307599399 |
| NFKBID | nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, delta [Source:HGNC Symbol;Acc:HGNC:15671] | signal transducer | 0 | 0 | 0.308275193 |
| CEBPA | CCAAT/enhancer binding protein (C/EBP), alpha [Source:HGNC Symbol;Acc:HGNC:1833] | transcription | 0 | 0 | 0.308795008 |
| PLXNA4 | plexin A4 [Source:HGNC Symbol;Acc:HGNC:9102] | receptor | 0 | 1 | 0.309299579 |
| PDE1B | phosphodiesterase 1B, calmodulin-dependent [Source:HGNC Symbol;Acc:HGNC:8775] | signal transducer | 1 | 0 | 0.309497536 |
| TOMM70A | translocase of outer mitochondrial membrane 70 homolog A (S. cerevisiae) [Source:HGNC Symbol;Acc:HGNC:11985] | adaptor | 0 | 1 | 0.30993982 |
| SOCS6 | suppressor of cytokine signaling 6 [Source:HGNC Symbol;Acc:HGNC:16833] | effector | 0 | 1 | 0.31049577 |
| YWHAB | tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, beta [Source:HGNC Symbol;Acc:HGNC:12849] | regulator | 1 | 0 | 0.310578434 |
| GRM5 | glutamate receptor, metabotropic 5 [Source:HGNC Symbol;Acc:HGNC:4597] | receptor | 0 | 0 | 0.310791299 |
| TP63 | tumor protein p63 [Source:HGNC Symbol;Acc:HGNC:15979] | transcription | 0 | 1 | 0.31086529 |
| PRKCQ | protein kinase C, theta [Source:HGNC Symbol;Acc:HGNC:9410] | signal transducer | 1 | 0 | 0.311738213 |
| TPST1 | tyrosylprotein sulfotransferase 1 [Source:HGNC Symbol;Acc:HGNC:12020] | UC | 0 | 1 | 0.311914786 |
| FYN | FYN proto-oncogene, Src family tyrosine kinase [Source:HGNC Symbol;Acc:HGNC:4037] | signal transducer | 1 | 0 | 0.312265614 |
| PRKCA | protein kinase C, alpha [Source:HGNC Symbol;Acc:HGNC:9393] | regulator | 1 | 1 | 0.312323191 |
| AKT2 | v-akt murine thymoma viral oncogene homolog 2 [Source:HGNC Symbol;Acc:HGNC:392] | signal transducer | 0 | 1 | 0.312380223 |
| GNAI2 | guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 2 [Source:HGNC Symbol;Acc:HGNC:4385] | regulator | 0 | 1 | 0.312737797 |
| ARF6 | ADP-ribosylation factor 6 [Source:HGNC Symbol;Acc:HGNC:659] | accessory molecule | 0 | 1 | 0.3131058 |
| RHOA | ras homolog family member A [Source:HGNC Symbol;Acc:HGNC:667] | signal transducer | 0 | 0 | 0.313435412 |
| ACTR2 | ARP2 actin-related protein 2 homolog (yeast) [Source:HGNC Symbol;Acc:HGNC:169] | UC | 1 | 0 | 0.313585384 |
| PRKCE | protein kinase C, epsilon [Source:HGNC Symbol;Acc:HGNC:9401] | signal transducer | 1 | 1 | 0.313592134 |
| OTUD7B | OTU deubiquitinase 7B [Source:HGNC Symbol;Acc:HGNC:16683] | regulator | 0 | 1 | 0.313797483 |

| | | | | | |
|---|---|---|---|---|---|
| FZD1 | frizzled class receptor 1 [Source:HGNC Symbol;Acc:HGNC:4038] | receptor | 0 | 1 | 0.313954651 |
| RAB11A | RAB11A, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:9760] | accessory molecule | 0 | 1 | 0.314017511 |
| THBS1 | thrombospondin 1 [Source:HGNC Symbol;Acc:HGNC:11785] | UC | 0 | 1 | 0.31417857 |
| PLEC | plectin [Source:HGNC Symbol;Acc:HGNC:9069] | UC | 0 | 1 | 0.314224851 |
| RPL19 | ribosomal protein L19 [Source:HGNC Symbol;Acc:HGNC:10312] | regulator | 0 | 1 | 0.314663909 |
| RASGRP1 | RAS guanyl releasing protein 1 (calcium and DAG-regulated) [Source:HGNC Symbol;Acc:HGNC:9878] | signal transducer | 1 | 0 | 0.314713246 |
| MAP3K7 | mitogen-activated protein kinase kinase kinase 7 [Source:HGNC Symbol;Acc:HGNC:6859] | signal transducer | 1 | 1 | 0.314772293 |
| PRKAR2B | protein kinase, cAMP-dependent, regulatory, type II, beta [Source:HGNC Symbol;Acc:HGNC:9392] | regulator | 1 | 0 | 0.315109458 |
| IRF4 | interferon regulatory factor 4 [Source:HGNC Symbol;Acc:HGNC:6119] | transcription | 1 | 1 | 0.31547986 |
| PTPN11 | protein tyrosine phosphatase, non-receptor type 11 [Source:HGNC Symbol;Acc:HGNC:9644] | regulator | 1 | 1 | 0.315698153 |
| HDAC11 | histone deacetylase 11 [Source:HGNC Symbol;Acc:HGNC:19086] | regulator | 0 | 1 | 0.315786777 |
| ERBB4 | erb-b2 receptor tyrosine kinase 4 [Source:HGNC Symbol;Acc:HGNC:3432] | receptor | 1 | 0 | 0.316406273 |
| PIK3CB | phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit beta [Source:HGNC Symbol;Acc:HGNC:8976] | signal transducer | 1 | 1 | 0.31664671 |
| FGF18 | fibroblast growth factor 18 [Source:HGNC Symbol;Acc:HGNC:3674] | effector | 1 | 0 | 0.316848936 |
| FGFR2 | fibroblast growth factor receptor 2 [Source:HGNC Symbol;Acc:HGNC:3689] | receptor | 1 | 0 | 0.317141933 |
| COPS5 | COP9 signalosome subunit 5 [Source:HGNC Symbol;Acc:HGNC:2240] | signal transducer | 0 | 1 | 0.317530373 |
| STUB1 | STIP1 homology and U-box containing protein 1, E3 ubiquitin protein ligase [Source:HGNC Symbol;Acc:HGNC:11427] | accessory molecule | 0 | 1 | 0.317577928 |
| FGF17 | fibroblast growth factor 17 [Source:HGNC Symbol;Acc:HGNC:3673] | effector | 1 | 0 | 0.317837097 |
| PELI3 | pellino E3 ubiquitin protein ligase family member 3 [Source:HGNC Symbol;Acc:HGNC:30010] | signal transducer | 0 | 1 | 0.317863843 |
| IFNG | interferon, gamma [Source:HGNC Symbol;Acc:HGNC:5438] | effector | 1 | 1 | 0.318543054 |
| MAPKAP1 | mitogen-activated protein kinase associated protein 1 [Source:HGNC Symbol;Acc:HGNC:18752] | signal transducer | 1 | 0 | 0.318642637 |
| RELB | v-rel avian reticuloendotheliosis viral oncogene homolog B [Source:HGNC Symbol;Acc:HGNC:9956] | transcription | 1 | 1 | 0.318660342 |
| NLRP3 | NLR family, pyrin domain containing 3 [Source:HGNC Symbol;Acc:HGNC:16400] | receptor | 1 | 1 | 0.31888745 |
| HSPA1A | heat shock 70kDa protein 1A [Source:HGNC Symbol;Acc:HGNC:5232] | effector | 0 | 1 | 0.319301373 |
| CDK6 | cyclin-dependent kinase 6 [Source:HGNC Symbol;Acc:HGNC:1777] | effector | 0 | 1 | 0.319420541 |

| | | | | | |
|---|---|---|---|---|---|
| VAV2 | vav 2 guanine nucleotide exchange factor [Source:HGNC Symbol;Acc:HGNC:12658] | signal transducer | 1 | 0 | 0.320119415 |
| RBCK1 | RanBP-type and C3HC4-type zinc finger containing 1 [Source:HGNC Symbol;Acc:HGNC:15864] | accessory molecule | 0 | 1 | 0.320167553 |
| BGN | biglycan [Source:HGNC Symbol;Acc:HGNC:1044] | UC | 0 | 1 | 0.320489563 |
| TUFM | Tu translation elongation factor, mitochondrial [Source:HGNC Symbol;Acc:HGNC:12420] | accessory molecule | 0 | 1 | 0.320565639 |
| ABCG1 | ATP-binding cassette, sub-family G (WHITE), member 1 [Source:HGNC Symbol;Acc:HGNC:73] | UC | 0 | 1 | 0.320857189 |
| MAPKAPK2 | mitogen-activated protein kinase-activated protein kinase 2 [Source:HGNC Symbol;Acc:HGNC:6887] | regulator | 1 | 1 | 0.321161018 |
| CCR7 | chemokine (C-C motif) receptor 7 [Source:HGNC Symbol;Acc:HGNC:1608] | secondary receptor | 0 | 1 | 0.321193279 |
| IMPDH2 | IMP (inosine 5'-monophosphate) dehydrogenase 2 [Source:HGNC Symbol;Acc:HGNC:6053] | UC | 0 | 1 | 0.321380305 |
| KPNA1 | karyopherin alpha 1 (importin alpha 5) [Source:HGNC Symbol;Acc:HGNC:6394] | accessory molecule | 0 | 1 | 0.321539263 |
| TBX21 | T-box 21 [Source:HGNC Symbol;Acc:HGNC:11599] | transcription | 0 | 0 | 0.321805077 |
| MAP2K6 | mitogen-activated protein kinase kinase 6 [Source:HGNC Symbol;Acc:HGNC:6846] | signal transducer | 1 | 1 | 0.321932708 |
| CALM1 | calmodulin 1 (phosphorylase kinase, delta) [Source:HGNC Symbol;Acc:HGNC:1442] | effector | 1 | 1 | 0.322249282 |
| KCNJ8 | potassium channel, inwardly rectifying subfamily J, member 8 [Source:HGNC Symbol;Acc:HGNC:6269] | UC | 0 | 1 | 0.322256342 |
| MAPK3 | mitogen-activated protein kinase 3 [Source:HGNC Symbol;Acc:HGNC:6877] | signal transducer | 1 | 0 | 0.322399938 |
| PCBP1 | poly(rC) binding protein 1 [Source:HGNC Symbol;Acc:HGNC:8647] | adaptor | 0 | 1 | 0.322973924 |
| KSR1 | kinase suppressor of ras 1 [Source:HGNC Symbol;Acc:HGNC:6465] | adaptor | 0 | 0 | 0.323816656 |
| CDC42 | cell division cycle 42 [Source:HGNC Symbol;Acc:HGNC:1736] | accessory molecule | 1 | 0 | 0.324102236 |
| RHEB | Ras homolog enriched in brain [Source:HGNC Symbol;Acc:HGNC:10011] | signal transducer | 0 | 1 | 0.324127367 |
| MDM2 | MDM2 proto-oncogene, E3 ubiquitin protein ligase [Source:HGNC Symbol;Acc:HGNC:6973] | regulator | 1 | 0 | 0.324167433 |
| FXR1 | fragile X mental retardation, autosomal homolog 1 [Source:HGNC Symbol;Acc:HGNC:4023] | receptor | 0 | 1 | 0.324659306 |
| DUSP7 | dual specificity phosphatase 7 [Source:HGNC Symbol;Acc:HGNC:3073] | regulator | 1 | 0 | 0.324776054 |
| NFKB2 | nuclear factor of kappa light polypeptide gene enhancer in B-cells 2 (p49/p100) [Source:HGNC Symbol;Acc:HGNC:7795] | transcription | 1 | 1 | 0.32500592 |
| POLR3D | polymerase (RNA) III (DNA directed) polypeptide D, 44kDa [Source:HGNC Symbol;Acc:HGNC:1080] | receptor | 1 | 0 | 0.325024856 |

| | | | | | |
|---|---|---|---|---|---|
| IRS2 | insulin receptor substrate 2 [Source:HGNC Symbol;Acc:HGNC:6126] | effector | 1 | 0 | 0.325544429 |
| POLR3A | polymerase (RNA) III (DNA directed) polypeptide A, 155kDa [Source:HGNC Symbol;Acc:HGNC:30074] | receptor | 1 | 0 | 0.325632758 |
| GPSM1 | G-protein signaling modulator 1 [Source:HGNC Symbol;Acc:HGNC:17858] | regulator | 0 | 1 | 0.32568234 |
| RPS6KA5 | ribosomal protein S6 kinase, 90kDa, polypeptide 5 [Source:HGNC Symbol;Acc:HGNC:10434] | signal transducer | 1 | 1 | 0.325821401 |
| TRIM56 | tripartite motif containing 56 [Source:HGNC Symbol;Acc:HGNC:19028] | regulator | 1 | 1 | 0.325983502 |
| IRF2BP1 | interferon regulatory factor 2 binding protein 1 [Source:HGNC Symbol;Acc:HGNC:21728] | transcription | 0 | 1 | 0.326137761 |
| MFN2 | mitofusin 2 [Source:HGNC Symbol;Acc:HGNC:16877] | accessory molecule | 0 | 1 | 0.326183575 |
| FGF7 | fibroblast growth factor 7 [Source:HGNC Symbol;Acc:HGNC:3685] | effector | 1 | 1 | 0.326231239 |
| WNT3A | wingless-type MMTV integration site family, member 3A [Source:HGNC Symbol;Acc:HGNC:15983] | effector | 0 | 1 | 0.326860104 |
| ARPC5 | actin related protein 2/3 complex, subunit 5, 16kDa [Source:HGNC Symbol;Acc:HGNC:708] | UC | 1 | 0 | 0.326891971 |
| ADARB1 | adenosine deaminase, RNA-specific, B1 [Source:HGNC Symbol;Acc:HGNC:226] | regulator | 1 | 0 | 0.327129262 |
| MAP2K4 | mitogen-activated protein kinase kinase 4 [Source:HGNC Symbol;Acc:HGNC:6844] | signal transducer | 1 | 1 | 0.3276245 |
| FGFR1 | fibroblast growth factor receptor 1 [Source:HGNC Symbol;Acc:HGNC:3688] | receptor | 1 | 0 | 0.327813138 |
| SIAH1 | siah E3 ubiquitin protein ligase 1 [Source:HGNC Symbol;Acc:HGNC:10857] | signal transducer | 0 | 1 | 0.328306803 |
| MAPK8 | mitogen-activated protein kinase 8 [Source:HGNC Symbol;Acc:HGNC:6881] | signal transducer | 1 | 1 | 0.328667573 |
| UBE2D2 | ubiquitin-conjugating enzyme E2D 2 [Source:HGNC Symbol;Acc:HGNC:12475] | regulator | 1 | 0 | 0.329008697 |
| C4A | complement component 4A (Rodgers blood group) [Source:HGNC Symbol;Acc:HGNC:1323] | signal transducer | 1 | 1 | 0.329399243 |
| RORA | RAR-related orphan receptor A [Source:HGNC Symbol;Acc:HGNC:10258] | transcription | 0 | 1 | 0.32943477 |
| PRKACA | protein kinase, cAMP-dependent, catalytic, alpha [Source:HGNC Symbol;Acc:HGNC:9380] | signal transducer | 1 | 1 | 0.329476177 |
| SIKE1 | suppressor of IKBKE 1 [Source:HGNC Symbol;Acc:HGNC:26119] | regulator | 1 | 0 | 0.329737517 |
| PIK3R1 | phosphoinositide-3-kinase, regulatory subunit 1 (alpha) [Source:HGNC Symbol;Acc:HGNC:8979] | signal transducer | 1 | 1 | 0.329886155 |
| CORO1A | coronin, actin binding protein, 1A [Source:HGNC Symbol;Acc:HGNC:2252] | accessory molecule | 1 | 0 | 0.329962424 |
| PTK2 | protein tyrosine kinase 2 [Source:HGNC Symbol;Acc:HGNC:9611] | regulator | 1 | 1 | 0.330632424 |
| ACKR4 | atypical chemokine receptor 4 [Source:HGNC Symbol;Acc:HGNC:1611] | secondary | 0 | 0 | 0.330835754 |

| | | | | | |
|---|---|---|---|---|---|
| | | receptor | | | |
| SP100 | SP100 nuclear antigen [Source:HGNC Symbol;Acc:HGNC:11206] | regulator | 1 | 0 | 0.331378705 |
| CFL1 | cofilin 1 (non-muscle) [Source:HGNC Symbol;Acc:HGNC:1874] | accessory molecule | 1 | 0 | 0.331648018 |
| NPTX1 | neuronal pentraxin I [Source:HGNC Symbol;Acc:HGNC:7952] | receptor | 0 | 0 | 0.332533466 |
| RAC2 | ras-related C3 botulinum toxin substrate 2 (rho family, small GTP binding protein Rac2) [Source:HGNC Symbol;Acc:HGNC:9802] | regulator | 0 | 1 | 0.332599114 |
| PDPK1 | 3-phosphoinositide dependent protein kinase 1 [Source:HGNC Symbol;Acc:HGNC:8816] | signal transducer | 1 | 0 | 0.332930697 |
| PRKACB | protein kinase, cAMP-dependent, catalytic, beta [Source:HGNC Symbol;Acc:HGNC:9381] | signal transducer | 1 | 0 | 0.333317975 |
| TRIM28 | tripartite motif containing 28 [Source:HGNC Symbol;Acc:HGNC:16384] | regulator | 1 | 1 | 0.333449228 |
| CYBB | cytochrome b-245, beta polypeptide [Source:HGNC Symbol;Acc:HGNC:2578] | effector | 1 | 1 | 0.333682113 |
| TGFB3 | transforming growth factor, beta 3 [Source:HGNC Symbol;Acc:HGNC:11769] | effector | 0 | 0 | 0.334210965 |
| FGFR3 | fibroblast growth factor receptor 3 [Source:HGNC Symbol;Acc:HGNC:3690] | receptor | 1 | 0 | 0.334251806 |
| LTB4R | leukotriene B4 receptor [Source:HGNC Symbol;Acc:HGNC:6713] | receptor | 0 | 1 | 0.33458054 |
| PDGFRA | platelet-derived growth factor receptor, alpha polypeptide [Source:HGNC Symbol;Acc:HGNC:8803] | secondary receptor | 1 | 0 | 0.334670444 |
| ZBTB1 | zinc finger and BTB domain containing 1 [Source:HGNC Symbol;Acc:HGNC:20259] | regulator | 1 | 0 | 0.335007298 |
| IRF6 | interferon regulatory factor 6 [Source:HGNC Symbol;Acc:HGNC:6121] | transcription | 1 | 1 | 0.335105575 |
| RCOR1 | REST corepressor 1 [Source:HGNC Symbol;Acc:HGNC:17441] | regulator | 0 | 1 | 0.335535815 |
| POLR3B | polymerase (RNA) III (DNA directed) polypeptide B [Source:HGNC Symbol;Acc:HGNC:30348] | receptor | 1 | 0 | 0.335761081 |
| CACTIN | cactin, spliceosome C complex subunit [Source:HGNC Symbol;Acc:HGNC:29938] | regulator | 1 | 0 | 0.335832424 |
| TRIM35 | tripartite motif containing 35 [Source:HGNC Symbol;Acc:HGNC:16285] | regulator | 1 | 0 | 0.33608181 |
| KDM4A | lysine (K)-specific demethylase 4A [Source:HGNC Symbol;Acc:HGNC:22978] | regulator | 0 | 1 | 0.336580095 |
| ADCY5 | adenylate cyclase 5 [Source:HGNC Symbol;Acc:HGNC:236] | UC | 1 | 0 | 0.336620884 |
| PIK3R4 | phosphoinositide-3-kinase, regulatory subunit 4 [Source:HGNC Symbol;Acc:HGNC:8982] | signal transducer | 1 | 0 | 0.33691866 |
| ITGB2 | integrin, beta 2 (complement component 3 receptor 3 and 4 subunit) [Source:HGNC Symbol;Acc:HGNC:6155] | receptor | 1 | 1 | 0.337128773 |
| RELA | v-rel avian reticuloendotheliosis viral oncogene homolog A [Source:HGNC Symbol;Acc:HGNC:9955] | transcription | 1 | 1 | 0.337325643 |
| PYDC1 | PYD (pyrin domain) containing 1 [Source:HGNC Symbol;Acc:HGNC:30261] | regulator | 1 | 0 | 0.337654163 |
| FOXO4 | forkhead box O4 [Source:HGNC Symbol;Acc:HGNC:7139] | transcription | 1 | 0 | 0.337853042 |

| | | | | | |
|---|---|---|---|---|---|
| IRF2 | interferon regulatory factor 2 [Source:HGNC Symbol;Acc:HGNC:6117] | transcription | 1 | 1 | 0.338475435 |
| TRIM62 | tripartite motif containing 62 [Source:HGNC Symbol;Acc:HGNC:25574] | regulator | 1 | 1 | 0.338595071 |
| PTGS2 | prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase) [Source:HGNC Symbol;Acc:HGNC:9605] | effector | 0 | 1 | 0.33903042 |
| CRK | v-crk avian sarcoma virus CT10 oncogene homolog [Source:HGNC Symbol;Acc:HGNC:2362] | adaptor | 1 | 0 | 0.339088196 |
| ADCY9 | adenylate cyclase 9 [Source:HGNC Symbol;Acc:HGNC:240] | UC | 1 | 0 | 0.339211687 |
| PDE1A | phosphodiesterase 1A, calmodulin-dependent [Source:HGNC Symbol;Acc:HGNC:8774] | signal transducer | 1 | 0 | 0.339223913 |
| RPS6KA1 | ribosomal protein S6 kinase, 90kDa, polypeptide 1 [Source:HGNC Symbol;Acc:HGNC:10430] | signal transducer | 1 | 0 | 0.339740667 |
| TRIM44 | tripartite motif containing 44 [Source:HGNC Symbol;Acc:HGNC:19016] | regulator | 0 | 0 | 0.340132397 |
| IRF8 | interferon regulatory factor 8 [Source:HGNC Symbol;Acc:HGNC:5358] | transcription | 1 | 1 | 0.340221042 |
| MYO10 | myosin X [Source:HGNC Symbol;Acc:HGNC:7593] | UC | 1 | 0 | 0.340332896 |
| ANXA2 | annexin A2 [Source:HGNC Symbol;Acc:HGNC:537] | UC | 0 | 1 | 0.340583585 |
| TCF4 | transcription factor 4 [Source:HGNC Symbol;Acc:HGNC:11634] | transcription | 0 | 1 | 0.34103592 |
| KRAS | Kirsten rat sarcoma viral oncogene homolog [Source:HGNC Symbol;Acc:HGNC:6407] | signal transducer | 1 | 0 | 0.3413167 |
| MIB1 | mindbomb E3 ubiquitin protein ligase 1 [Source:HGNC Symbol;Acc:HGNC:21086] | regulator | 0 | 0 | 0.341346481 |
| CAPZA2 | capping protein (actin filament) muscle Z-line, alpha 2 [Source:HGNC Symbol;Acc:HGNC:1490] | UC | 1 | 0 | 0.341528255 |
| PIAS2 | protein inhibitor of activated STAT, 2 [Source:HGNC Symbol;Acc:HGNC:17311] | regulator | 0 | 1 | 0.342215187 |
| YES1 | YES proto-oncogene 1, Src family tyrosine kinase [Source:HGNC Symbol;Acc:HGNC:12841] | signal transducer | 1 | 0 | 0.342223773 |
| ADCY8 | adenylate cyclase 8 (brain) [Source:HGNC Symbol;Acc:HGNC:239] | UC | 1 | 0 | 0.342247722 |
| MARCH5 | membrane-associated ring finger (C3HC4) 5 [Source:HGNC Symbol;Acc:HGNC:26025] | regulator | 0 | 1 | 0.342628218 |
| TCEB1 | transcription elongation factor B (SIII), polypeptide 1 (15kDa, elongin C) [Source:HGNC Symbol;Acc:HGNC:11617] | transcription | 0 | 1 | 0.342947699 |
| SOS1 | son of sevenless homolog 1 (Drosophila) [Source:HGNC Symbol;Acc:HGNC:11187] | regulator | 1 | 0 | 0.343000331 |
| UBE2D1 | ubiquitin-conjugating enzyme E2D 1 [Source:HGNC Symbol;Acc:HGNC:12474] | regulator | 1 | 0 | 0.343297756 |
| CXCR3 | chemokine (C-X-C motif) receptor 3 [Source:HGNC Symbol;Acc:HGNC:4540] | secondary receptor | 0 | 1 | 0.343322935 |
| RAF1 | Raf-1 proto-oncogene, serine/threonine kinase [Source:HGNC Symbol;Acc:HGNC:9829] | signal transducer | 1 | 0 | 0.344785278 |
| CEBPD | CCAAT/enhancer binding protein (C/EBP), delta [Source:HGNC Symbol;Acc:HGNC:1835] | transcription | 0 | 1 | 0.344961346 |

| | | | | | |
|---|---|---|---|---|---|
| CHUK | conserved helix-loop-helix ubiquitous kinase [Source:HGNC Symbol;Acc:HGNC:1974] | signal transducer | 1 | 1 | 0.345009055 |
| APP | amyloid beta (A4) precursor protein [Source:HGNC Symbol;Acc:HGNC:620] | UC | 1 | 0 | 0.345420616 |
| EP300 | E1A binding protein p300 [Source:HGNC Symbol;Acc:HGNC:3373] | transcription | 1 | 0 | 0.345427651 |
| MAPKAPK3 | mitogen-activated protein kinase-activated protein kinase 3 [Source:HGNC Symbol;Acc:HGNC:6888] | regulator | 1 | 0 | 0.345504884 |
| NFKBIA | nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha [Source:HGNC Symbol;Acc:HGNC:7797] | signal transducer | 1 | 1 | 0.345674792 |
| HRAS | Harvey rat sarcoma viral oncogene homolog [Source:HGNC Symbol;Acc:HGNC:5173] | adaptor | 1 | 1 | 0.345982876 |
| RPS27A | ribosomal protein S27a [Source:HGNC Symbol;Acc:HGNC:10417] | accessory molecule | 1 | 0 | 0.346041109 |
| CYP27B1 | cytochrome P450, family 27, subfamily B, polypeptide 1 [Source:HGNC Symbol;Acc:HGNC:2606] | UC | 1 | 0 | 0.346169703 |
| PURA | purine-rich element binding protein A [Source:HGNC Symbol;Acc:HGNC:9701] | transcription | 0 | 1 | 0.347438334 |
| GJA1 | gap junction protein, alpha 1, 43kDa [Source:HGNC Symbol;Acc:HGNC:4274] | UC | 0 | 1 | 0.347725838 |
| CCR4 | chemokine (C-C motif) receptor 4 [Source:HGNC Symbol;Acc:HGNC:1605] | secondary receptor | 0 | 1 | 0.348035011 |
| PRMT1 | protein arginine methyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:5187] | regulator | 0 | 1 | 0.348448236 |
| PELI2 | pellino E3 ubiquitin protein ligase family member 2 [Source:HGNC Symbol;Acc:HGNC:8828] | signal transducer | 1 | 1 | 0.348475847 |
| POLR3F | polymerase (RNA) III (DNA directed) polypeptide F, 39 kDa [Source:HGNC Symbol;Acc:HGNC:15763] | receptor | 1 | 0 | 0.349135534 |
| PTCH1 | patched 1 [Source:HGNC Symbol;Acc:HGNC:9585] | signal transducer | 0 | 1 | 0.349669686 |
| TRIM72 | tripartite motif containing 72, E3 ubiquitin protein ligase [Source:HGNC Symbol;Acc:HGNC:32671] | regulator | 0 | 0 | 0.349920454 |
| LCP2 | lymphocyte cytosolic protein 2 (SH2 domain containing leukocyte protein of 76kDa) [Source:HGNC Symbol;Acc:HGNC:6529] | signal transducer | 1 | 0 | 0.350061965 |
| GNB2L1 | guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1 [Source:HGNC Symbol;Acc:HGNC:4399] | regulator | 0 | 1 | 0.350475147 |
| CTLA4 | cytotoxic T-lymphocyte-associated protein 4 [Source:HGNC Symbol;Acc:HGNC:2505] | regulator | 0 | 1 | 0.350651385 |
| OLR1 | oxidized low density lipoprotein (lectin-like) receptor 1 [Source:HGNC Symbol;Acc:HGNC:8133] | receptor | 0 | 0 | 0.350761407 |
| NFAT5 | nuclear factor of activated T-cells 5, tonicity-responsive [Source:HGNC Symbol;Acc:HGNC:7774] | transcription | 0 | 1 | 0.351064864 |
| MEF2A | myocyte enhancer factor 2A [Source:HGNC Symbol;Acc:HGNC:6993] | transcription | 1 | 0 | 0.351170767 |
| GCH1 | GTP cyclohydrolase 1 [Source:HGNC Symbol;Acc:HGNC:4193] | regulator | 1 | 0 | 0.351800333 |
| ARPC2 | actin related protein 2/3 complex, subunit 2, 34kDa [Source:HGNC Symbol;Acc:HGNC:705] | UC | 1 | 0 | 0.352169225 |
| SNX27 | sorting nexin family member 27 [Source:HGNC Symbol;Acc:HGNC:20073] | accessory | 0 | 1 | 0.352212122 |

| | | | | | |
|---|---|---|---|---|---|
| | | molecule | | | |
| DNM1L | dynamin 1-like [Source:HGNC Symbol;Acc:HGNC:2973] | effector | 0 | 1 | 0.352377047 |
| TSC22D3 | TSC22 domain family, member 3 [Source:HGNC Symbol;Acc:HGNC:3051] | regulator | 0 | 1 | 0.352786985 |
| CAMKK2 | calcium/calmodulin-dependent protein kinase kinase 2, beta [Source:HGNC Symbol;Acc:HGNC:1470] | signal transducer | 0 | 1 | 0.352799615 |
| SPI1 | Spi-1 proto-oncogene [Source:HGNC Symbol;Acc:HGNC:11241] | transcription | 0 | 1 | 0.352998699 |
| TRIM23 | tripartite motif containing 23 [Source:HGNC Symbol;Acc:HGNC:660] | signal transducer | 0 | 1 | 0.353174683 |
| PKN1 | protein kinase N1 [Source:HGNC Symbol;Acc:HGNC:9405] | signal transducer | 0 | 1 | 0.353220703 |
| UBE2D3 | ubiquitin-conjugating enzyme E2D 3 [Source:HGNC Symbol;Acc:HGNC:12476] | regulator | 1 | 0 | 0.353228558 |
| ELK1 | ELK1, member of ETS oncogene family [Source:HGNC Symbol;Acc:HGNC:3321] | transcription | 1 | 0 | 0.353269421 |
| WASL | Wiskott-Aldrich syndrome-like [Source:HGNC Symbol;Acc:HGNC:12735] | UC | 1 | 0 | 0.353345588 |
| PPARG | peroxisome proliferator-activated receptor gamma [Source:HGNC Symbol;Acc:HGNC:9236] | receptor | 1 | 1 | 0.353489488 |
| ASGR2 | asialoglycoprotein receptor 2 [Source:HGNC Symbol;Acc:HGNC:743] | receptor | 0 | 0 | 0.353757682 |
| NFIL3 | nuclear factor, interleukin 3 regulated [Source:HGNC Symbol;Acc:HGNC:7787] | transcription | 0 | 1 | 0.354060707 |
| AGO4 | argonaute RISC catalytic component 4 [Source:HGNC Symbol;Acc:HGNC:18424] | effector | 1 | 0 | 0.354062187 |
| DEAF1 | DEAF1 transcription factor [Source:HGNC Symbol;Acc:HGNC:14677] | regulator | 0 | 0 | 0.354239644 |
| KRT1 | keratin 1 [Source:HGNC Symbol;Acc:HGNC:6412] | regulator | 1 | 0 | 0.354488564 |
| ADAR | adenosine deaminase, RNA-specific [Source:HGNC Symbol;Acc:HGNC:225] | regulator | 1 | 1 | 0.35458221 |
| IL21 | interleukin 21 [Source:HGNC Symbol;Acc:HGNC:6005] | effector | 0 | 1 | 0.354607828 |
| WDR34 | WD repeat domain 34 [Source:HGNC Symbol;Acc:HGNC:28296] | regulator | 0 | 1 | 0.354726121 |
| FER | fer (fps/fes related) tyrosine kinase [Source:HGNC Symbol;Acc:HGNC:3655] | signal transducer | 1 | 0 | 0.354903648 |
| NKIRAS2 | NFKB inhibitor interacting Ras-like 2 [Source:HGNC Symbol;Acc:HGNC:17898] | signal transducer | 0 | 1 | 0.35516482 |
| DAPK3 | death-associated protein kinase 3 [Source:HGNC Symbol;Acc:HGNC:2676] | signal transducer | 1 | 0 | 0.356481095 |
| SEMA3A | sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3A [Source:HGNC Symbol;Acc:HGNC:10723] | effector | 0 | 1 | 0.356662583 |
| CCR6 | chemokine (C-C motif) receptor 6 [Source:HGNC Symbol;Acc:HGNC:1607] | secondary receptor | 1 | 1 | 0.356841774 |
| TP73 | tumor protein p73 [Source:HGNC Symbol;Acc:HGNC:12003] | transcription | 0 | 1 | 0.356995793 |

| | | | | | |
|---|---|---|---|---|---|
| SUMO1 | small ubiquitin-like modifier 1 [Source:HGNC Symbol;Acc:HGNC:12502] | accessory molecule | 1 | 0 | 0.357339524 |
| CXCL11 | chemokine (C-X-C motif) ligand 11 [Source:HGNC Symbol;Acc:HGNC:10638] | effector | 0 | 1 | 0.35739984 |
| ANGPT1 | angiopoietin 1 [Source:HGNC Symbol;Acc:HGNC:484] | regulator | 0 | 1 | 0.357514105 |
| TSC2 | tuberous sclerosis 2 [Source:HGNC Symbol;Acc:HGNC:12363] | regulator | 1 | 0 | 0.357693116 |
| RICTOR | RPTOR independent companion of MTOR, complex 2 [Source:HGNC Symbol;Acc:HGNC:28611] | accessory molecule | 1 | 1 | 0.357713327 |
| MAP1LC3C | microtubule-associated protein 1 light chain 3 gamma [Source:HGNC Symbol;Acc:HGNC:13353] | effector | 0 | 1 | 0.358057922 |
| RPS6KA4 | ribosomal protein S6 kinase, 90kDa, polypeptide 4 [Source:HGNC Symbol;Acc:HGNC:10433] | signal transducer | 0 | 1 | 0.358314465 |
| RCAN1 | regulator of calcineurin 1 [Source:HGNC Symbol;Acc:HGNC:3040] | regulator | 0 | 1 | 0.35833646 |
| ANXA1 | annexin A1 [Source:HGNC Symbol;Acc:HGNC:533] | UC | 0 | 1 | 0.358537725 |
| RNF41 | ring finger protein 41, E3 ubiquitin protein ligase [Source:HGNC Symbol;Acc:HGNC:18401] | regulator | 0 | 1 | 0.358651382 |
| PIK3R5 | phosphoinositide-3-kinase, regulatory subunit 5 [Source:HGNC Symbol;Acc:HGNC:30035] | signal transducer | 0 | 0 | 0.358717934 |
| MAPK9 | mitogen-activated protein kinase 9 [Source:HGNC Symbol;Acc:HGNC:6886] | signal transducer | 1 | 1 | 0.358873007 |
| CRKL | v-crk avian sarcoma virus CT10 oncogene homolog-like [Source:HGNC Symbol;Acc:HGNC:2363] | adaptor | 0 | 1 | 0.359178774 |
| ADCY2 | adenylate cyclase 2 (brain) [Source:HGNC Symbol;Acc:HGNC:233] | UC | 1 | 0 | 0.359798903 |
| TRIM17 | tripartite motif containing 17 [Source:HGNC Symbol;Acc:HGNC:13430] | regulator | 0 | 0 | 0.359918513 |
| HIST1H2BK | histone cluster 1, H2bk [Source:HGNC Symbol;Acc:HGNC:13954] | UC | 1 | 0 | 0.359975666 |
| TRIM24 | tripartite motif containing 24 [Source:HGNC Symbol;Acc:HGNC:11812] | regulator | 0 | 1 | 0.360294107 |
| ABL1 | ABL proto-oncogene 1, non-receptor tyrosine kinase [Source:HGNC Symbol;Acc:HGNC:76] | signal transducer | 1 | 1 | 0.360587413 |
| CTSD | cathepsin D [Source:HGNC Symbol;Acc:HGNC:2529] | accessory molecule | 0 | 1 | 0.360890763 |
| BIRC3 | baculoviral IAP repeat containing 3 [Source:HGNC Symbol;Acc:HGNC:591] | regulator | 1 | 1 | 0.361157098 |
| F2R | coagulation factor II (thrombin) receptor [Source:HGNC Symbol;Acc:HGNC:3537] | secondary receptor | 0 | 1 | 0.361181706 |
| RUNX3 | runt-related transcription factor 3 [Source:HGNC Symbol;Acc:HGNC:10473] | transcription | 0 | 1 | 0.361732653 |
| IP6K1 | inositol hexakisphosphate kinase 1 [Source:HGNC Symbol;Acc:HGNC:18360] | signal transducer | 0 | 1 | 0.361768605 |
| DHX29 | DEAH (Asp-Glu-Ala-His) box polypeptide 29 [Source:HGNC Symbol;Acc:HGNC:15815] | receptor | 0 | 0 | 0.361989167 |

| | | | | | |
|---|---|---|---|---|---|
| CAMK4 | calcium/calmodulin-dependent protein kinase IV [Source:HGNC Symbol;Acc:HGNC:1464] | signal transducer | 1 | 0 | 0.362323423 |
| BCL3 | B-cell CLL/lymphoma 3 [Source:HGNC Symbol;Acc:HGNC:998] | signal transducer | 0 | 1 | 0.362331751 |
| DLK1 | delta-like 1 homolog (Drosophila) [Source:HGNC Symbol;Acc:HGNC:2907] | UC | 0 | 1 | 0.362335649 |
| CARD11 | caspase recruitment domain family, member 11 [Source:HGNC Symbol;Acc:HGNC:16393] | signal transducer | 1 | 0 | 0.362591929 |
| TLR9 | toll-like receptor 9 [Source:HGNC Symbol;Acc:HGNC:15633] | receptor | 1 | 1 | 0.362609968 |
| FCGR2B | Fc fragment of IgG, low affinity IIb, receptor (CD32) [Source:HGNC Symbol;Acc:HGNC:3618] | receptor | 0 | 0 | 0.363120227 |
| ZFPM1 | zinc finger protein, FOG family member 1 [Source:HGNC Symbol;Acc:HGNC:19762] | regulator | 0 | 1 | 0.363148266 |
| ATF1 | activating transcription factor 1 [Source:HGNC Symbol;Acc:HGNC:783] | transcription | 1 | 0 | 0.363366146 |
| DAB2IP | DAB2 interacting protein [Source:HGNC Symbol;Acc:HGNC:17294] | regulator | 1 | 1 | 0.363431516 |
| NKIRAS1 | NFKB inhibitor interacting Ras-like 1 [Source:HGNC Symbol;Acc:HGNC:17899] | signal transducer | 0 | 1 | 0.363480978 |
| IFITM3 | interferon induced transmembrane protein 3 [Source:HGNC Symbol;Acc:HGNC:5414] | effector | 1 | 1 | 0.363681703 |
| TRIM61 | tripartite motif containing 61 [Source:HGNC Symbol;Acc:HGNC:24339] | regulator | 0 | 1 | 0.364180802 |
| COLEC11 | collectin sub-family member 11 [Source:HGNC Symbol;Acc:HGNC:17213] | receptor | 0 | 0 | 0.364316899 |
| C3 | complement component 3 [Source:HGNC Symbol;Acc:HGNC:1318] | signal transducer | 1 | 1 | 0.364706511 |
| DRD2 | dopamine receptor D2 [Source:HGNC Symbol;Acc:HGNC:3023] | UC | 0 | 0 | 0.364753967 |
| TGFB2 | transforming growth factor, beta 2 [Source:HGNC Symbol;Acc:HGNC:11768] | effector | 0 | 0 | 0.365073361 |
| GNAI3 | guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 3 [Source:HGNC Symbol;Acc:HGNC:4387] | signal transducer | 0 | 1 | 0.366192941 |
| S100B | S100 calcium binding protein B [Source:HGNC Symbol;Acc:HGNC:10500] | effector | 1 | 0 | 0.366271005 |
| SCAMP5 | secretory carrier membrane protein 5 [Source:HGNC Symbol;Acc:HGNC:30386] | UC | 0 | 1 | 0.366272153 |
| IKBKE | inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase epsilon [Source:HGNC Symbol;Acc:HGNC:14552] | signal transducer | 1 | 1 | 0.366500654 |
| CITED1 | Cbp/p300-interacting transactivator, with Glu/Asp-rich carboxy-terminal domain, 1 [Source:HGNC Symbol;Acc:HGNC:1986] | regulator | 1 | 0 | 0.366525325 |
| ITCH | itchy E3 ubiquitin protein ligase [Source:HGNC Symbol;Acc:HGNC:13890] | regulator | 1 | 1 | 0.367342806 |
| S100A8 | S100 calcium binding protein A8 [Source:HGNC Symbol;Acc:HGNC:10498] | effector | 1 | 1 | 0.367373333 |
| CD28 | CD28 molecule [Source:HGNC Symbol;Acc:HGNC:1653] | receptor | 1 | 0 | 0.367450833 |
| B2M | beta-2-microglobulin [Source:HGNC Symbol;Acc:HGNC:914] | UC | 1 | 0 | 0.368247834 |

| | | | | | |
|---|---|---|---|---|---|
| DDX1 | DEAD (Asp-Glu-Ala-Asp) box helicase 1 [Source:HGNC Symbol;Acc:HGNC:2734] | receptor | 0 | 1 | 0.36877805 |
| BECN1 | beclin 1, autophagy related [Source:HGNC Symbol;Acc:HGNC:1034] | effector | 0 | 1 | 0.368809908 |
| TGFB1 | transforming growth factor, beta 1 [Source:HGNC Symbol;Acc:HGNC:11766] | effector | 1 | 0 | 0.368911523 |
| DICER1 | dicer 1, ribonuclease type III [Source:HGNC Symbol;Acc:HGNC:17098] | UC | 0 | 1 | 0.368941348 |
| FES | FES proto-oncogene, tyrosine kinase [Source:HGNC Symbol;Acc:HGNC:3657] | regulator | 1 | 0 | 0.368971279 |
| UBQLN1 | ubiquilin 1 [Source:HGNC Symbol;Acc:HGNC:12508] | regulator | 0 | 1 | 0.36909733 |
| BCL2 | B-cell CLL/lymphoma 2 [Source:HGNC Symbol;Acc:HGNC:990] | signal transducer | 1 | 1 | 0.369450419 |
| XBP1 | X-box binding protein 1 [Source:HGNC Symbol;Acc:HGNC:12801] | transcription | 0 | 1 | 0.369638733 |
| PIN1 | peptidylprolyl cis/trans isomerase, NIMA-interacting 1 [Source:HGNC Symbol;Acc:HGNC:8988] | regulator | 1 | 1 | 0.369808208 |
| DHX36 | DEAH (Asp-Glu-Ala-His) box polypeptide 36 [Source:HGNC Symbol;Acc:HGNC:14410] | receptor | 1 | 1 | 0.369811264 |
| ATG5 | autophagy related 5 [Source:HGNC Symbol;Acc:HGNC:589] | regulator | 1 | 1 | 0.370219091 |
| LGALS1 | lectin, galactoside-binding, soluble, 1 [Source:HGNC Symbol;Acc:HGNC:6561] | receptor | 0 | 1 | 0.371391444 |
| CAMK2D | calcium/calmodulin-dependent protein kinase II delta [Source:HGNC Symbol;Acc:HGNC:1462] | signal transducer | 1 | 0 | 0.371623336 |
| SOCS1 | suppressor of cytokine signaling 1 [Source:HGNC Symbol;Acc:HGNC:19383] | effector | 1 | 1 | 0.371703611 |
| RNF5 | ring finger protein 5, E3 ubiquitin protein ligase [Source:HGNC Symbol;Acc:HGNC:10068] | regulator | 0 | 1 | 0.372098028 |
| MATK | megakaryocyte-associated tyrosine kinase [Source:HGNC Symbol;Acc:HGNC:6906] | signal transducer | 1 | 0 | 0.372109608 |
| TRAF2 | TNF receptor-associated factor 2 [Source:HGNC Symbol;Acc:HGNC:12032] | signal transducer | 1 | 1 | 0.372509467 |
| POLR1D | polymerase (RNA) I polypeptide D, 16kDa [Source:HGNC Symbol;Acc:HGNC:20422] | receptor | 1 | 0 | 0.372619345 |
| FBXO9 | F-box protein 9 [Source:HGNC Symbol;Acc:HGNC:13588] | regulator | 1 | 0 | 0.372674791 |
| CASP8 | caspase 8, apoptosis-related cysteine peptidase [Source:HGNC Symbol;Acc:HGNC:1509] | signal transducer | 1 | 1 | 0.373663858 |
| ILF3 | interleukin enhancer binding factor 3, 90kDa [Source:HGNC Symbol;Acc:HGNC:6038] | regulator | 0 | 1 | 0.374070353 |
| HMGB3 | high mobility group box 3 [Source:HGNC Symbol;Acc:HGNC:5004] | receptor | 0 | 1 | 0.374194878 |
| IL18 | interleukin 18 [Source:HGNC Symbol;Acc:HGNC:5986] | effector | 0 | 1 | 0.374229262 |
| SENP2 | SUMO1/sentrin/SMT3 specific peptidase 2 [Source:HGNC Symbol;Acc:HGNC:23116] | regulator | 0 | 1 | 0.374355028 |
| NRAS | neuroblastoma RAS viral (v-ras) oncogene homolog [Source:HGNC Symbol;Acc:HGNC:7989] | signal transducer | 1 | 1 | 0.374771674 |
| C1QC | complement component 1, q subcomponent, C chain [Source:HGNC Symbol;Acc:HGNC:1245] | receptor | 1 | 1 | 0.374829667 |
| GSK3A | glycogen synthase kinase 3 alpha [Source:HGNC Symbol;Acc:HGNC:4616] | signal | 1 | 0 | 0.374893341 |

| | | | | | |
|---|---|---|---|---|---|
| SNCA | synuclein, alpha (non A4 component of amyloid precursor) [Source:HGNC Symbol;Acc:HGNC:11138] | transducer effector | 1 | 1 | 0.375124088 |
| DDIT3 | DNA-damage-inducible transcript 3 [Source:HGNC Symbol;Acc:HGNC:2726] | effector | 0 | 1 | 0.375794936 |
| BAD | BCL2-associated agonist of cell death [Source:HGNC Symbol;Acc:HGNC:936] | signal transducer | 1 | 0 | 0.3760722 |
| USP18 | ubiquitin specific peptidase 18 [Source:HGNC Symbol;Acc:HGNC:12616] | regulator | 1 | 0 | 0.376557392 |
| VAMP2 | vesicle-associated membrane protein 2 (synaptobrevin 2) [Source:HGNC Symbol;Acc:HGNC:12643] | accessory molecule | 1 | 0 | 0.377160941 |
| WAS | Wiskott-Aldrich syndrome [Source:HGNC Symbol;Acc:HGNC:12731] | signal transducer | 1 | 1 | 0.377407335 |
| STYK1 | serine/threonine/tyrosine kinase 1 [Source:HGNC Symbol;Acc:HGNC:18889] | regulator | 1 | 0 | 0.377530329 |
| CD274 | CD274 molecule [Source:HGNC Symbol;Acc:HGNC:17635] | effector | 0 | 1 | 0.377600737 |
| TMED7 | transmembrane emp24 protein transport domain containing 7 [Source:HGNC Symbol;Acc:HGNC:24253] | UC | 0 | 0 | 0.378048481 |
| PLCG2 | phospholipase C, gamma 2 (phosphatidylinositol-specific) [Source:HGNC Symbol;Acc:HGNC:9066] | signal transducer | 1 | 1 | 0.378316612 |
| TRIM47 | tripartite motif containing 47 [Source:HGNC Symbol;Acc:HGNC:19020] | regulator | 0 | 1 | 0.378428795 |
| CTSL | cathepsin L [Source:HGNC Symbol;Acc:HGNC:2537] | accessory molecule | 1 | 0 | 0.378491464 |
| IRF5 | interferon regulatory factor 5 [Source:HGNC Symbol;Acc:HGNC:6120] | transcription | 1 | 1 | 0.378588373 |
| PDE1C | phosphodiesterase 1C, calmodulin-dependent 70kDa [Source:HGNC Symbol;Acc:HGNC:8776] | signal transducer | 1 | 0 | 0.378670648 |
| HIF1AN | hypoxia inducible factor 1, alpha subunit inhibitor [Source:HGNC Symbol;Acc:HGNC:17113] | regulator | 0 | 1 | 0.378786868 |
| CD53 | CD53 molecule [Source:HGNC Symbol;Acc:HGNC:1686] | regulator | 0 | 1 | 0.378864054 |
| CLEC16A | C-type lectin domain family 16, member A [Source:HGNC Symbol;Acc:HGNC:29013] | receptor | 0 | 0 | 0.379064458 |
| CXCL9 | chemokine (C-X-C motif) ligand 9 [Source:HGNC Symbol;Acc:HGNC:7098] | effector | 0 | 1 | 0.379150167 |
| LIMK1 | LIM domain kinase 1 [Source:HGNC Symbol;Acc:HGNC:6613] | signal transducer | 1 | 0 | 0.379452157 |
| PTPN1 | protein tyrosine phosphatase, non-receptor type 1 [Source:HGNC Symbol;Acc:HGNC:9642] | regulator | 1 | 0 | 0.379686441 |
| DUSP3 | dual specificity phosphatase 3 [Source:HGNC Symbol;Acc:HGNC:3069] | regulator | 1 | 0 | 0.379834033 |
| MAP3K8 | mitogen-activated protein kinase kinase kinase 8 [Source:HGNC Symbol;Acc:HGNC:6860] | signal transducer | 0 | 1 | 0.379845898 |
| FGB | fibrinogen beta chain [Source:HGNC Symbol;Acc:HGNC:3662] | effector | 1 | 0 | 0.380155231 |

| | | | | | |
|---|---|---|---|---|---|
| HMGB1 | high mobility group box 1 [Source:HGNC Symbol;Acc:HGNC:4983] | receptor | 1 | 1 | 0.380300297 |
| FOXO1 | forkhead box O1 [Source:HGNC Symbol;Acc:HGNC:3819] | transcription | 1 | 0 | 0.380433155 |
| HP | haptoglobin [Source:HGNC Symbol;Acc:HGNC:5141] | UC | 0 | 1 | 0.380761606 |
| TRAIP | TRAF interacting protein [Source:HGNC Symbol;Acc:HGNC:30764] | regulator | 0 | 1 | 0.381145335 |
| PSMA7 | proteasome (prosome, macropain) subunit, alpha type, 7 [Source:HGNC Symbol;Acc:HGNC:9536] | accessory molecule | 0 | 1 | 0.381227023 |
| DEFB123 | defensin, beta 123 [Source:HGNC Symbol;Acc:HGNC:18103] | effector | 1 | 0 | 0.381344133 |
| HIST1H2BC | histone cluster 1, H2bc [Source:HGNC Symbol;Acc:HGNC:4757] | UC | 1 | 0 | 0.381620742 |
| JAK2 | Janus kinase 2 [Source:HGNC Symbol;Acc:HGNC:6192] | signal transducer | 1 | 1 | 0.381644823 |
| GRAP2 | GRB2-related adaptor protein 2 [Source:HGNC Symbol;Acc:HGNC:4563] | adaptor | 1 | 0 | 0.3818419 |
| TRIM50 | tripartite motif containing 50 [Source:HGNC Symbol;Acc:HGNC:19017] | regulator | 0 | 1 | 0.382008664 |
| UBE2V1 | ubiquitin-conjugating enzyme E2 variant 1 [Source:HGNC Symbol;Acc:HGNC:12494] | signal transducer | 1 | 0 | 0.382099112 |
| STIM1 | stromal interaction molecule 1 [Source:HGNC Symbol;Acc:HGNC:11386] | UC | 0 | 1 | 0.382221928 |
| DEFB106B | defensin, beta 106B [Source:HGNC Symbol;Acc:HGNC:28879] | effector | 0 | 0 | 0.382277393 |
| NUMBL | numb homolog (Drosophila)-like [Source:HGNC Symbol;Acc:HGNC:8061] | signal transducer | 0 | 1 | 0.382507991 |
| IP6K2 | inositol hexakisphosphate kinase 2 [Source:HGNC Symbol;Acc:HGNC:17313] | signal transducer | 1 | 0 | 0.38263946 |
| BCL2L1 | BCL2-like 1 [Source:HGNC Symbol;Acc:HGNC:992] | regulator | 1 | 0 | 0.382761067 |
| ARHGEF2 | Rho/Rac guanine nucleotide exchange factor (GEF) 2 [Source:HGNC Symbol;Acc:HGNC:682] | receptor | 1 | 0 | 0.38276771 |
| FLT4 | fms-related tyrosine kinase 4 [Source:HGNC Symbol;Acc:HGNC:3767] | secondary receptor | 0 | 1 | 0.382875406 |
| DHCR24 | 24-dehydrocholesterol reductase [Source:HGNC Symbol;Acc:HGNC:2859] | effector | 0 | 1 | 0.383014176 |
| GABARAP | GABA(A) receptor-associated protein [Source:HGNC Symbol;Acc:HGNC:4067] | accessory molecule | 0 | 0 | 0.383539468 |
| SUGT1 | SGT1, suppressor of G2 allele of SKP1 (S. cerevisiae) [Source:HGNC Symbol;Acc:HGNC:16987] | regulator | 1 | 1 | 0.383987396 |
| POLR2F | polymerase (RNA) II (DNA directed) polypeptide F [Source:HGNC Symbol;Acc:HGNC:9193] | transcription | 1 | 1 | 0.384158315 |
| DDOST | dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit (non-catalytic) [Source:HGNC Symbol;Acc:HGNC:2728] | UC | 1 | 0 | 0.384464098 |
| EIF2AK2 | eukaryotic translation initiation factor 2-alpha kinase 2 [Source:HGNC Symbol;Acc:HGNC:9437] | receptor | 1 | 1 | 0.384579179 |
| EGR1 | early growth response 1 [Source:HGNC Symbol;Acc:HGNC:3238] | effector | 1 | 1 | 0.38518347 |

| | | | | | |
|---|---|---|---|---|---|
| JAK3 | Janus kinase 3 [Source:HGNC Symbol;Acc:HGNC:6193] | signal transducer | 1 | 1 | 0.385227948 |
| TRIM11 | tripartite motif containing 11 [Source:HGNC Symbol;Acc:HGNC:16281] | regulator | 1 | 0 | 0.385246302 |
| PACSIN1 | protein kinase C and casein kinase substrate in neurons 1 [Source:HGNC Symbol;Acc:HGNC:8570] | UC | 0 | 1 | 0.385413333 |
| NUB1 | negative regulator of ubiquitin-like proteins 1 [Source:HGNC Symbol;Acc:HGNC:17623] | regulator | 1 | 0 | 0.385460427 |
| IL13RA1 | interleukin 13 receptor, alpha 1 [Source:HGNC Symbol;Acc:HGNC:5974] | secondary receptor | 0 | 1 | 0.385483596 |
| TAB1 | TGF-beta activated kinase 1/MAP3K7 binding protein 1 [Source:HGNC Symbol;Acc:HGNC:18157] | signal transducer | 1 | 0 | 0.386492215 |
| ARPC4 | actin related protein 2/3 complex, subunit 4, 20kDa [Source:HGNC Symbol;Acc:HGNC:707] | UC | 1 | 0 | 0.386613718 |
| SAMHD1 | SAM domain and HD domain 1 [Source:HGNC Symbol;Acc:HGNC:15925] | receptor | 1 | 1 | 0.386684258 |
| DEFA6 | defensin, alpha 6, Paneth cell-specific [Source:HGNC Symbol;Acc:HGNC:2765] | effector | 1 | 1 | 0.386780602 |
| WIPF1 | WAS/WASL interacting protein family, member 1 [Source:HGNC Symbol;Acc:HGNC:12736] | UC | 1 | 0 | 0.386908495 |
| CD59 | CD59 molecule, complement regulatory protein [Source:HGNC Symbol;Acc:HGNC:1689] | regulator | 1 | 0 | 0.387114197 |
| CFB | complement factor B [Source:HGNC Symbol;Acc:HGNC:1037] | regulator | 1 | 1 | 0.387686256 |
| PML | promyelocytic leukemia [Source:HGNC Symbol;Acc:HGNC:9113] | regulator | 1 | 1 | 0.388058551 |
| S100A10 | S100 calcium binding protein A10 [Source:HGNC Symbol;Acc:HGNC:10487] | effector | 0 | 1 | 0.388059848 |
| RPL39 | ribosomal protein L39 [Source:HGNC Symbol;Acc:HGNC:10350] | UC | 1 | 0 | 0.388263834 |
| NAIP | NLR family, apoptosis inhibitory protein [Source:HGNC Symbol;Acc:HGNC:7634] | receptor | 1 | 1 | 0.388362024 |
| TRIM8 | tripartite motif containing 8 [Source:HGNC Symbol;Acc:HGNC:15579] | regulator | 1 | 1 | 0.38911488 |
| TRAF6 | TNF receptor-associated factor 6, E3 ubiquitin protein ligase [Source:HGNC Symbol;Acc:HGNC:12036] | signal transducer | 1 | 1 | 0.389340931 |
| SCARB1 | scavenger receptor class B, member 1 [Source:HGNC Symbol;Acc:HGNC:1664] | receptor | 0 | 1 | 0.38993657 |
| TRIM63 | tripartite motif containing 63, E3 ubiquitin protein ligase [Source:HGNC Symbol;Acc:HGNC:16007] | regulator | 0 | 1 | 0.39071142 |
| KITLG | KIT ligand [Source:HGNC Symbol;Acc:HGNC:6343] | effector | 1 | 1 | 0.390931668 |
| ERBB2 | erb-b2 receptor tyrosine kinase 2 [Source:HGNC Symbol;Acc:HGNC:3430] | receptor | 1 | 0 | 0.390984156 |
| VAV3 | vav 3 guanine nucleotide exchange factor [Source:HGNC Symbol;Acc:HGNC:12659] | signal transducer | 1 | 0 | 0.390987394 |
| FGF9 | fibroblast growth factor 9 [Source:HGNC Symbol;Acc:HGNC:3687] | effector | 1 | 0 | 0.391050269 |
| NPTX2 | neuronal pentraxin II [Source:HGNC Symbol;Acc:HGNC:7953] | receptor | 0 | 0 | 0.391088404 |
| DDX60 | DEAD (Asp-Glu-Ala-Asp) box polypeptide 60 [Source:HGNC Symbol;Acc:HGNC:25942] | receptor | 1 | 0 | 0.392067041 |
| GLRX | glutaredoxin (thioltransferase) [Source:HGNC Symbol;Acc:HGNC:4330] | signal transducer | 0 | 1 | 0.392243091 |

| | | | | | |
|---|---|---|---|---|---|
| TRIM4 | tripartite motif containing 4 [Source:HGNC Symbol;Acc:HGNC:16275] | regulator | 1 | 0 | 0.392354894 |
| POLR3GL | polymerase (RNA) III (DNA directed) polypeptide G (32kD)-like [Source:HGNC Symbol;Acc:HGNC:28466] | receptor | 1 | 0 | 0.392435754 |
| HDAC2 | histone deacetylase 2 [Source:HGNC Symbol;Acc:HGNC:4853] | regulator | 0 | 1 | 0.392491475 |
| VAMP7 | vesicle-associated membrane protein 7 [Source:HGNC Symbol;Acc:HGNC:11486] | accessory molecule | 1 | 0 | 0.392807787 |
| IFNGR1 | interferon gamma receptor 1 [Source:HGNC Symbol;Acc:HGNC:5439] | secondary receptor | 1 | 1 | 0.393574677 |
| NFATC1 | nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 1 [Source:HGNC Symbol;Acc:HGNC:7775] | transcription | 1 | 0 | 0.393765135 |
| RASGEF1B | RasGEF domain family, member 1B [Source:HGNC Symbol;Acc:HGNC:24881] | effector | 0 | 1 | 0.393926102 |
| ARG1 | arginase 1 [Source:HGNC Symbol;Acc:HGNC:663] | regulator | 0 | 1 | 0.394015893 |
| RORC | RAR-related orphan receptor C [Source:HGNC Symbol;Acc:HGNC:10260] | transcription | 0 | 0 | 0.39414676 |
| TSC1 | tuberous sclerosis 1 [Source:HGNC Symbol;Acc:HGNC:12362] | regulator | 0 | 1 | 0.394330266 |
| RAD23A | RAD23 homolog A (S. cerevisiae) [Source:HGNC Symbol;Acc:HGNC:9812] | regulator | 0 | 1 | 0.394622587 |
| NR3C1 | nuclear receptor subfamily 3, group C, member 1 (glucocorticoid receptor) [Source:HGNC Symbol;Acc:HGNC:7978] | secondary receptor | 0 | 1 | 0.395642979 |
| PPBP | pro-platelet basic protein (chemokine (C-X-C motif) ligand 7) [Source:HGNC Symbol;Acc:HGNC:9240] | effector | 0 | 0 | 0.395663231 |
| FOS | FBJ murine osteosarcoma viral oncogene homolog [Source:HGNC Symbol;Acc:HGNC:3796] | transcription | 1 | 0 | 0.396241718 |
| FGF16 | fibroblast growth factor 16 [Source:HGNC Symbol;Acc:HGNC:3672] | effector | 1 | 0 | 0.396421404 |
| HOXA9 | homeobox A9 [Source:HGNC Symbol;Acc:HGNC:5109] | transcription | 0 | 1 | 0.397229093 |
| PPAPDC1A | phosphatidic acid phosphatase type 2 domain containing 1A [Source:HGNC Symbol;Acc:HGNC:23531] | UC | 1 | 0 | 0.397443347 |
| FBXW5 | F-box and WD repeat domain containing 5 [Source:HGNC Symbol;Acc:HGNC:13613] | regulator | 0 | 1 | 0.397931757 |
| DCN | decorin [Source:HGNC Symbol;Acc:HGNC:2705] | UC | 0 | 1 | 0.398173272 |
| STAR | steroidogenic acute regulatory protein [Source:HGNC Symbol;Acc:HGNC:11359] | accessory molecule | 1 | 0 | 0.3984246 |
| CLEC11A | C-type lectin domain family 11, member A [Source:HGNC Symbol;Acc:HGNC:10576] | receptor | 0 | 0 | 0.398670263 |
| ASCC3 | activating signal cointegrator 1 complex subunit 3 [Source:HGNC Symbol;Acc:HGNC:18697] | regulator | 0 | 1 | 0.399229837 |
| SH2D1A | SH2 domain containing 1A [Source:HGNC Symbol;Acc:HGNC:10820] | regulator | 0 | 1 | 0.399395902 |
| DEFB4B | defensin, beta 4B [Source:HGNC Symbol;Acc:HGNC:30193] | effector | 0 | 0 | 0.399491002 |
| NR1H3 | nuclear receptor subfamily 1, group H, member 3 [Source:HGNC Symbol;Acc:HGNC:7966] | transcription | 0 | 1 | 0.399496022 |

| | | | | | |
|---|---|---|---|---|---|
| IFNE | interferon, epsilon [Source:HGNC Symbol;Acc:HGNC:18163] | effector | 1 | 0 | 0.399775613 |
| MYH2 | myosin, heavy chain 2, skeletal muscle, adult [Source:HGNC Symbol;Acc:HGNC:7572] | UC | 1 | 0 | 0.399997083 |
| BID | BH3 interacting domain death agonist [Source:HGNC Symbol;Acc:HGNC:1050] | regulator | 0 | 1 | 0.400231186 |
| GATA6 | GATA binding protein 6 [Source:HGNC Symbol;Acc:HGNC:4174] | transcription | 0 | 1 | 0.400376682 |
| TMSB4X | thymosin beta 4, X-linked [Source:HGNC Symbol;Acc:HGNC:11881] | effector | 0 | 0 | 0.400499143 |
| LST1 | leukocyte specific transcript 1 [Source:HGNC Symbol;Acc:HGNC:14189] | signal transducer | 0 | 1 | 0.400957646 |
| FOXO3 | forkhead box O3 [Source:HGNC Symbol;Acc:HGNC:3821] | regulator | 1 | 1 | 0.401083343 |
| LCK | LCK proto-oncogene, Src family tyrosine kinase [Source:HGNC Symbol;Acc:HGNC:6524] | signal transducer | 1 | 0 | 0.401134075 |
| CFLAR | CASP8 and FADD-like apoptosis regulator [Source:HGNC Symbol;Acc:HGNC:1876] | regulator | 0 | 1 | 0.401258026 |
| PDGFRB | platelet-derived growth factor receptor, beta polypeptide [Source:HGNC Symbol;Acc:HGNC:8804] | secondary receptor | 1 | 0 | 0.401348007 |
| HSPBP1 | HSPA (heat shock 70kDa) binding protein, cytoplasmic cochaperone 1 [Source:HGNC Symbol;Acc:HGNC:24989] | regulator | 0 | 1 | 0.401376158 |
| IL23A | interleukin 23, alpha subunit p19 [Source:HGNC Symbol;Acc:HGNC:15488] | effector | 1 | 1 | 0.401609337 |
| KLRG1 | killer cell lectin-like receptor subfamily G, member 1 [Source:HGNC Symbol;Acc:HGNC:6380] | receptor | 1 | 0 | 0.401736045 |
| ADCY7 | adenylate cyclase 7 [Source:HGNC Symbol;Acc:HGNC:238] | UC | 1 | 0 | 0.40184175 |
| FLI1 | Fli-1 proto-oncogene, ETS transcription factor [Source:HGNC Symbol;Acc:HGNC:3749] | transcription | 0 | 1 | 0.40186861 |
| MX1 | MX dynamin-like GTPase 1 [Source:HGNC Symbol;Acc:HGNC:7532] | effector | 1 | 1 | 0.402341859 |
| CLEC3B | C-type lectin domain family 3, member B [Source:HGNC Symbol;Acc:HGNC:11891] | receptor | 0 | 0 | 0.402415979 |
| ABCF1 | ATP-binding cassette, sub-family F (GCN20), member 1 [Source:HGNC Symbol;Acc:HGNC:70] | receptor | 0 | 0 | 0.402819532 |
| TNRC6C | trinucleotide repeat containing 6C [Source:HGNC Symbol;Acc:HGNC:29318] | regulator | 1 | 0 | 0.402979934 |
| RIPK1 | receptor (TNFRSF)-interacting serine-threonine kinase 1 [Source:HGNC Symbol;Acc:HGNC:10019] | signal transducer | 1 | 1 | 0.403407766 |
| NPR2 | natriuretic peptide receptor 2 [Source:HGNC Symbol;Acc:HGNC:7944] | receptor | 0 | 0 | 0.403585832 |
| TRIM46 | tripartite motif containing 46 [Source:HGNC Symbol;Acc:HGNC:19019] | regulator | 0 | 0 | 0.403618572 |
| ADRB2 | adrenoceptor beta 2, surface [Source:HGNC Symbol;Acc:HGNC:286] | UC | 0 | 1 | 0.403712409 |
| DEFB103B | defensin, beta 103B [Source:HGNC Symbol;Acc:HGNC:31702] | effector | 0 | 1 | 0.403803076 |
| PRKCSH | protein kinase C substrate 80K-H [Source:HGNC Symbol;Acc:HGNC:9411] | UC | 1 | 0 | 0.403928987 |
| ARPC3 | actin related protein 2/3 complex, subunit 3, 21kDa [Source:HGNC Symbol;Acc:HGNC:706] | UC | 1 | 0 | 0.404249738 |
| TNFRSF1A | tumor necrosis factor receptor superfamily, member 1A [Source:HGNC Symbol;Acc:HGNC:11916] | secondary receptor | 0 | 1 | 0.404349834 |

| | | | | | |
|---|---|---|---|---|---|
| NOS2 | nitric oxide synthase 2, inducible [Source:HGNC Symbol;Acc:HGNC:7873] | effector | 1 | 1 | 0.404959173 |
| S100A9 | S100 calcium binding protein A9 [Source:HGNC Symbol;Acc:HGNC:10499] | effector | 1 | 1 | 0.405302683 |
| ARHGAP15 | Rho GTPase activating protein 15 [Source:HGNC Symbol;Acc:HGNC:21030] | regulator | 0 | 1 | 0.405666092 |
| UNC93B1 | unc-93 homolog B1 (C. elegans) [Source:HGNC Symbol;Acc:HGNC:13481] | accessory molecule | 1 | 1 | 0.405816624 |
| DUSP1 | dual specificity phosphatase 1 [Source:HGNC Symbol;Acc:HGNC:3064] | regulator | 0 | 1 | 0.405943805 |
| LGALS4 | lectin, galactoside-binding, soluble, 4 [Source:HGNC Symbol;Acc:HGNC:6565] | receptor | 0 | 1 | 0.406101234 |
| CLEC2L | C-type lectin domain family 2, member L [Source:HGNC Symbol;Acc:HGNC:21969] | receptor | 0 | 0 | 0.406116863 |
| CCL5 | chemokine (C-C motif) ligand 5 [Source:HGNC Symbol;Acc:HGNC:10632] | effector | 1 | 1 | 0.407393035 |
| CNOT8 | CCR4-NOT transcription complex, subunit 8 [Source:HGNC Symbol;Acc:HGNC:9207] | regulator | 0 | 1 | 0.407441593 |
| LTBR | lymphotoxin beta receptor (TNFR superfamily, member 3) [Source:HGNC Symbol;Acc:HGNC:6718] | secondary receptor | 0 | 1 | 0.407564295 |
| IFNA1 | interferon, alpha 1 [Source:HGNC Symbol;Acc:HGNC:5417] | effector | 1 | 1 | 0.407573844 |
| ATG16L1 | autophagy related 16-like 1 (S. cerevisiae) [Source:HGNC Symbol;Acc:HGNC:21498] | regulator | 0 | 1 | 0.407762727 |
| TNFSF10 | tumor necrosis factor (ligand) superfamily, member 10 [Source:HGNC Symbol;Acc:HGNC:11925] | effector | 0 | 1 | 0.408314111 |
| PTGES2 | prostaglandin E synthase 2 [Source:HGNC Symbol;Acc:HGNC:17822] | UC | 0 | 1 | 0.408522843 |
| GAB1 | GRB2-associated binding protein 1 [Source:HGNC Symbol;Acc:HGNC:4066] | adaptor | 1 | 1 | 0.410024111 |
| HIST1H2BF | histone cluster 1, H2bf [Source:HGNC Symbol;Acc:HGNC:4752] | UC | 0 | 0 | 0.410128839 |
| NOS3 | nitric oxide synthase 3 (endothelial cell) [Source:HGNC Symbol;Acc:HGNC:7876] | effector | 0 | 0 | 0.410286183 |
| OPTN | optineurin [Source:HGNC Symbol;Acc:HGNC:17142] | receptor | 0 | 1 | 0.410373027 |
| APOBEC3A | apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3A [Source:HGNC Symbol;Acc:HGNC:17343] | effector | 1 | 1 | 0.410457313 |
| POLR2L | polymerase (RNA) II (DNA directed) polypeptide L, 7.6kDa [Source:HGNC Symbol;Acc:HGNC:9199] | receptor | 1 | 0 | 0.410672696 |
| TRIM3 | tripartite motif containing 3 [Source:HGNC Symbol;Acc:HGNC:10064] | regulator | 0 | 0 | 0.410714885 |
| HSP90B1 | heat shock protein 90kDa beta (Grp94), member 1 [Source:HGNC Symbol;Acc:HGNC:12028] | accessory molecule | 1 | 1 | 0.410781567 |
| TECPR1 | tectonin beta-propeller repeat containing 1 [Source:HGNC Symbol;Acc:HGNC:22214] | signal transducer | 0 | 1 | 0.411110369 |
| PSMB8 | proteasome (prosome, macropain) subunit, beta type, 8 [Source:HGNC Symbol;Acc:HGNC:9545] | effector | 1 | 0 | 0.412156583 |
| MAP3K1 | mitogen-activated protein kinase kinase kinase 1, E3 ubiquitin protein ligase [Source:HGNC Symbol;Acc:HGNC:6848] | signal transducer | 1 | 1 | 0.412394807 |
| IL2RG | interleukin 2 receptor, gamma [Source:HGNC Symbol;Acc:HGNC:6010] | secondary receptor | 0 | 1 | 0.412545485 |

| | | | | | |
|---|---|---|---|---|---|
| TRIM37 | tripartite motif containing 37 [Source:HGNC Symbol;Acc:HGNC:7523] | regulator | 0 | 1 | 0.412828556 |
| FGF1 | fibroblast growth factor 1 (acidic) [Source:HGNC Symbol;Acc:HGNC:3665] | effector | 1 | 0 | 0.41299939 |
| AKAP10 | A kinase (PRKA) anchor protein 10 [Source:HGNC Symbol;Acc:HGNC:368] | signal transducer | 0 | 1 | 0.412999793 |
| IL15 | interleukin 15 [Source:HGNC Symbol;Acc:HGNC:5977] | effector | 0 | 1 | 0.413265668 |
| MAP3K4 | mitogen-activated protein kinase kinase kinase 4 [Source:HGNC Symbol;Acc:HGNC:6856] | signal transducer | 0 | 1 | 0.413267599 |
| SOCS3 | suppressor of cytokine signaling 3 [Source:HGNC Symbol;Acc:HGNC:19391] | effector | 1 | 1 | 0.413667604 |
| CCL2 | chemokine (C-C motif) ligand 2 [Source:HGNC Symbol;Acc:HGNC:10618] | effector | 1 | 1 | 0.413882327 |
| NCK1 | NCK adaptor protein 1 [Source:HGNC Symbol;Acc:HGNC:7664] | adaptor | 1 | 0 | 0.414978475 |
| CCNA2 | cyclin A2 [Source:HGNC Symbol;Acc:HGNC:1578] | regulator | 0 | 1 | 0.415077251 |
| ABL2 | ABL proto-oncogene 2, non-receptor tyrosine kinase [Source:HGNC Symbol;Acc:HGNC:77] | signal transducer | 1 | 0 | 0.415091352 |
| IFNA8 | interferon, alpha 8 [Source:HGNC Symbol;Acc:HGNC:5429] | effector | 1 | 0 | 0.41527635 |
| MAP3K5 | mitogen-activated protein kinase kinase kinase 5 [Source:HGNC Symbol;Acc:HGNC:6857] | signal transducer | 1 | 1 | 0.415987468 |
| IL10 | interleukin 10 [Source:HGNC Symbol;Acc:HGNC:5962] | effector | 0 | 1 | 0.416129693 |
| GRK5 | G protein-coupled receptor kinase 5 [Source:HGNC Symbol;Acc:HGNC:4544] | regulator | 0 | 1 | 0.416385013 |
| MALT1 | MALT1 paracaspase [Source:HGNC Symbol;Acc:HGNC:6819] | signal transducer | 1 | 1 | 0.416493916 |
| GATA4 | GATA binding protein 4 [Source:HGNC Symbol;Acc:HGNC:4173] | transcription | 0 | 1 | 0.416567057 |
| IFI6 | interferon, alpha-inducible protein 6 [Source:HGNC Symbol;Acc:HGNC:4054] | effector | 1 | 1 | 0.416584297 |
| TLR3 | toll-like receptor 3 [Source:HGNC Symbol;Acc:HGNC:11849] | receptor | 1 | 1 | 0.417296456 |
| TNFAIP3 | tumor necrosis factor, alpha-induced protein 3 [Source:HGNC Symbol;Acc:HGNC:11896] | regulator | 1 | 1 | 0.41770691 |
| UBE2L6 | ubiquitin-conjugating enzyme E2L 6 [Source:HGNC Symbol;Acc:HGNC:12490] | regulator | 1 | 0 | 0.417755804 |
| STAT6 | signal transducer and activator of transcription 6, interleukin-4 induced [Source:HGNC Symbol;Acc:HGNC:11368] | transcription | 1 | 1 | 0.418131858 |
| XRCC5 | X-ray repair complementing defective repair in Chinese hamster cells 5 (double-strand-break rejoining) [Source:HGNC Symbol;Acc:HGNC:12833] | receptor | 1 | 1 | 0.418237996 |
| IFNA13 | interferon, alpha 13 [Source:HGNC Symbol;Acc:HGNC:5419] | effector | 0 | 0 | 0.419026957 |
| MKNK1 | MAP kinase interacting serine/threonine kinase 1 [Source:HGNC Symbol;Acc:HGNC:7110] | signal transducer | 0 | 1 | 0.419391295 |
| NR1I2 | nuclear receptor subfamily 1, group I, member 2 [Source:HGNC Symbol;Acc:HGNC:7968] | transcription | 0 | 1 | 0.419470141 |

| | | | | | |
|---|---|---|---|---|---|
| AQP3 | aquaporin 3 (Gill blood group) [Source:HGNC Symbol;Acc:HGNC:636] | UC | 0 | 1 | 0.419501486 |
| ITK | IL2-inducible T-cell kinase [Source:HGNC Symbol;Acc:HGNC:6171] | signal transducer | 1 | 0 | 0.420501173 |
| AIF1 | allograft inflammatory factor 1 [Source:HGNC Symbol;Acc:HGNC:352] | UC | 1 | 0 | 0.420666657 |
| PPP2R5B | protein phosphatase 2, regulatory subunit B', beta [Source:HGNC Symbol;Acc:HGNC:9310] | signal transducer | 0 | 0 | 0.420728124 |
| FAU | Finkel-Biskis-Reilly murine sarcoma virus (FBR-MuSV) ubiquitously expressed [Source:HGNC Symbol;Acc:HGNC:3597] | UC | 1 | 0 | 0.420954932 |
| OLFM4 | olfactomedin 4 [Source:HGNC Symbol;Acc:HGNC:17190] | regulator | 0 | 1 | 0.421306364 |
| CD8A | CD8a molecule [Source:HGNC Symbol;Acc:HGNC:1706] | UC | 0 | 1 | 0.42155269 |
| ARPC1A | actin related protein 2/3 complex, subunit 1A, 41kDa [Source:HGNC Symbol;Acc:HGNC:703] | UC | 1 | 0 | 0.421608957 |
| ADCY4 | adenylate cyclase 4 [Source:HGNC Symbol;Acc:HGNC:235] | UC | 1 | 0 | 0.421699939 |
| IRAK1 | interleukin-1 receptor-associated kinase 1 [Source:HGNC Symbol;Acc:HGNC:6112] | signal transducer | 1 | 1 | 0.421886032 |
| DEFB104B | defensin, beta 104B [Source:HGNC Symbol;Acc:HGNC:26165] | effector | 0 | 0 | 0.421939129 |
| SIRT1 | sirtuin 1 [Source:HGNC Symbol;Acc:HGNC:14929] | UC | 0 | 1 | 0.421957278 |
| IFITM1 | interferon induced transmembrane protein 1 [Source:HGNC Symbol;Acc:HGNC:5412] | effector | 1 | 1 | 0.422526891 |
| POLR2K | polymerase (RNA) II (DNA directed) polypeptide K, 7.0kDa [Source:HGNC Symbol;Acc:HGNC:9198] | receptor | 1 | 0 | 0.422739138 |
| TRIM16 | tripartite motif containing 16 [Source:HGNC Symbol;Acc:HGNC:17241] | regulator | 0 | 0 | 0.4228631 |
| LRRFIP1 | leucine rich repeat (in FLII) interacting protein 1 [Source:HGNC Symbol;Acc:HGNC:6702] | receptor | 1 | 1 | 0.422872738 |
| PRKACG | protein kinase, cAMP-dependent, catalytic, gamma [Source:HGNC Symbol;Acc:HGNC:9382] | signal transducer | 1 | 0 | 0.422938456 |
| KAT2B | K(lysine) acetyltransferase 2B [Source:HGNC Symbol;Acc:HGNC:8638] | UC | 0 | 1 | 0.422951437 |
| MRC2 | mannose receptor, C type 2 [Source:HGNC Symbol;Acc:HGNC:16875] | receptor | 0 | 0 | 0.423233288 |
| POLR2E | polymerase (RNA) II (DNA directed) polypeptide E, 25kDa [Source:HGNC Symbol;Acc:HGNC:9192] | receptor | 1 | 0 | 0.423566774 |
| SYP | synaptophysin [Source:HGNC Symbol;Acc:HGNC:11506] | signal transducer | 0 | 1 | 0.423658673 |
| VEGFC | vascular endothelial growth factor C [Source:HGNC Symbol;Acc:HGNC:12682] | effector | 0 | 1 | 0.423784521 |
| PIK3AP1 | phosphoinositide-3-kinase adaptor protein 1 [Source:HGNC Symbol;Acc:HGNC:30034] | signal transducer | 0 | 1 | 0.424185357 |
| NFKB1 | nuclear factor of kappa light polypeptide gene enhancer in B-cells 1 [Source:HGNC Symbol;Acc:HGNC:7794] | transcription | 1 | 1 | 0.42430376 |
| KLRK1 | killer cell lectin-like receptor subfamily K, member 1 [Source:HGNC Symbol;Acc:HGNC:18788] | receptor | 1 | 0 | 0.424445569 |

| | | | | | |
|---|---|---|---|---|---|
| CIITA | class II, major histocompatibility complex, transactivator [Source:HGNC Symbol;Acc:HGNC:7067] | transcription | 1 | 0 | 0.424896536 |
| CISH | cytokine inducible SH2-containing protein [Source:HGNC Symbol;Acc:HGNC:1984] | effector | 0 | 1 | 0.425239477 |
| AKT1S1 | AKT1 substrate 1 (proline-rich) [Source:HGNC Symbol;Acc:HGNC:28426] | signal transducer | 1 | 0 | 0.425682806 |
| EDIL3 | EGF-like repeats and discoidin I-like domains 3 [Source:HGNC Symbol;Acc:HGNC:3173] | UC | 0 | 1 | 0.425843681 |
| HSF1 | heat shock transcription factor 1 [Source:HGNC Symbol;Acc:HGNC:5224] | transcription | 0 | 1 | 0.425951627 |
| IL12A | interleukin 12A [Source:HGNC Symbol;Acc:HGNC:5969] | effector | 0 | 1 | 0.42738224 |
| NLRP4 | NLR family, pyrin domain containing 4 [Source:HGNC Symbol;Acc:HGNC:22943] | receptor | 1 | 1 | 0.427429865 |
| DUSP16 | dual specificity phosphatase 16 [Source:HGNC Symbol;Acc:HGNC:17909] | regulator | 0 | 1 | 0.427650682 |
| CAV1 | caveolin 1, caveolae protein, 22kDa [Source:HGNC Symbol;Acc:HGNC:1527] | accessory molecule | 0 | 1 | 0.428849086 |
| MCOLN1 | mucolipin 1 [Source:HGNC Symbol;Acc:HGNC:13356] | UC | 0 | 0 | 0.428910677 |
| SFTPD | surfactant protein D [Source:HGNC Symbol;Acc:HGNC:10803] | receptor | 1 | 1 | 0.429002657 |
| TNFSF11 | tumor necrosis factor (ligand) superfamily, member 11 [Source:HGNC Symbol;Acc:HGNC:11926] | effector | 0 | 1 | 0.429011887 |
| AAMP | angio-associated, migratory cell protein [Source:HGNC Symbol;Acc:HGNC:18] | regulator | 0 | 1 | 0.429033132 |
| LGMN | legumain [Source:HGNC Symbol;Acc:HGNC:9472] | accessory molecule | 1 | 1 | 0.429068205 |
| TRIM21 | tripartite motif containing 21 [Source:HGNC Symbol;Acc:HGNC:11312] | regulator | 1 | 1 | 0.429713117 |
| CCNT1 | cyclin T1 [Source:HGNC Symbol;Acc:HGNC:1599] | regulator | 0 | 1 | 0.42972286 |
| SELK | NA | UC | 0 | 1 | 0.429871156 |
| BCL2A1 | BCL2-related protein A1 [Source:HGNC Symbol;Acc:HGNC:991] | regulator | 0 | 1 | 0.429944843 |
| SPON2 | spondin 2, extracellular matrix protein [Source:HGNC Symbol;Acc:HGNC:11253] | receptor | 1 | 1 | 0.429956007 |
| SQSTM1 | sequestosome 1 [Source:HGNC Symbol;Acc:HGNC:11280] | receptor | 0 | 1 | 0.430182377 |
| ABCA1 | ATP-binding cassette, sub-family A (ABC1), member 1 [Source:HGNC Symbol;Acc:HGNC:29] | UC | 0 | 1 | 0.430421329 |
| HIF1A | hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor) [Source:HGNC Symbol;Acc:HGNC:4910] | transcription | 0 | 1 | 0.430974247 |
| PRKDC | protein kinase, DNA-activated, catalytic polypeptide [Source:HGNC Symbol;Acc:HGNC:9413] | signal transducer | 1 | 1 | 0.430999384 |
| MYO1C | myosin IC [Source:HGNC Symbol;Acc:HGNC:7597] | UC | 1 | 0 | 0.431269279 |
| WNT2B | wingless-type MMTV integration site family, member 2B [Source:HGNC Symbol;Acc:HGNC:12781] | effector | 0 | 1 | 0.431296827 |
| CLEC19A | C-type lectin domain family 19, member A [Source:HGNC Symbol;Acc:HGNC:34522] | receptor | 0 | 0 | 0.431488928 |
| RFTN1 | raftlin, lipid raft linker 1 [Source:HGNC Symbol;Acc:HGNC:30278] | receptor | 0 | 1 | 0.431616648 |

| | | | | | |
|---|---|---|---|---|---|
| FGF4 | fibroblast growth factor 4 [Source:HGNC Symbol;Acc:HGNC:3682] | effector | 1 | 0 | 0.431741303 |
| UCHL1 | ubiquitin carboxyl-terminal esterase L1 (ubiquitin thiolesterase) [Source:HGNC Symbol;Acc:HGNC:12513] | regulator | 0 | 1 | 0.431843353 |
| SFTPA2 | surfactant protein A2 [Source:HGNC Symbol;Acc:HGNC:10799] | receptor | 0 | 1 | 0.432035809 |
| CSF1R | colony stimulating factor 1 receptor [Source:HGNC Symbol;Acc:HGNC:2433] | secondary receptor | 1 | 0 | 0.432206014 |
| POLR3H | polymerase (RNA) III (DNA directed) polypeptide H (22.9kD) [Source:HGNC Symbol;Acc:HGNC:30349] | receptor | 1 | 0 | 0.432381226 |
| LEP | leptin [Source:HGNC Symbol;Acc:HGNC:6553] | UC | 0 | 1 | 0.432796533 |
| ULK1 | unc-51 like autophagy activating kinase 1 [Source:HGNC Symbol;Acc:HGNC:12558] | signal transducer | 0 | 0 | 0.432860543 |
| TACR1 | tachykinin receptor 1 [Source:HGNC Symbol;Acc:HGNC:11526] | secondary receptor | 0 | 0 | 0.432950203 |
| ASS1 | argininosuccinate synthase 1 [Source:HGNC Symbol;Acc:HGNC:758] | UC | 1 | 0 | 0.433298557 |
| NCF4 | neutrophil cytosolic factor 4, 40kDa [Source:HGNC Symbol;Acc:HGNC:7662] | effector | 0 | 0 | 0.433709894 |
| HIST1H2BI | histone cluster 1, H2bi [Source:HGNC Symbol;Acc:HGNC:4756] | UC | 1 | 0 | 0.43388837 |
| SLC26A6 | solute carrier family 26 (anion exchanger), member 6 [Source:HGNC Symbol;Acc:HGNC:14472] | UC | 1 | 0 | 0.433946838 |
| EPS8 | epidermal growth factor receptor pathway substrate 8 [Source:HGNC Symbol;Acc:HGNC:3420] | regulator | 0 | 1 | 0.433954397 |
| SLAMF7 | SLAM family member 7 [Source:HGNC Symbol;Acc:HGNC:21394] | receptor | 1 | 1 | 0.433957749 |
| CD3G | CD3g molecule, gamma (CD3-TCR complex) [Source:HGNC Symbol;Acc:HGNC:1675] | receptor | 1 | 0 | 0.434475337 |
| FGF8 | fibroblast growth factor 8 (androgen-induced) [Source:HGNC Symbol;Acc:HGNC:3686] | effector | 1 | 0 | 0.434539272 |
| FRS2 | fibroblast growth factor receptor substrate 2 [Source:HGNC Symbol;Acc:HGNC:16971] | adaptor | 1 | 0 | 0.434963626 |
| HIST2H2BE | histone cluster 2, H2be [Source:HGNC Symbol;Acc:HGNC:4760] | UC | 1 | 0 | 0.435364347 |
| CASP7 | caspase 7, apoptosis-related cysteine peptidase [Source:HGNC Symbol;Acc:HGNC:1508] | signal transducer | 0 | 1 | 0.435378847 |
| NR4A1 | nuclear receptor subfamily 4, group A, member 1 [Source:HGNC Symbol;Acc:HGNC:7980] | receptor | 1 | 0 | 0.435382673 |
| RIPK2 | receptor-interacting serine-threonine kinase 2 [Source:HGNC Symbol;Acc:HGNC:10020] | adaptor | 1 | 1 | 0.435854912 |
| PDGFA | platelet-derived growth factor alpha polypeptide [Source:HGNC Symbol;Acc:HGNC:8799] | effector | 1 | 0 | 0.435890651 |
| PHLPP1 | PH domain and leucine rich repeat protein phosphatase 1 [Source:HGNC Symbol;Acc:HGNC:20610] | signal transducer | 1 | 0 | 0.43628641 |
| CRCP | CGRP receptor component [Source:HGNC Symbol;Acc:HGNC:17888] | receptor | 1 | 0 | 0.436294915 |
| KYNU | kynureninase [Source:HGNC Symbol;Acc:HGNC:6469] | UC | 1 | 0 | 0.436348509 |
| TAX1BP1 | Tax1 (human T-cell leukemia virus type I) binding protein 1 [Source:HGNC Symbol;Acc:HGNC:11575] | regulator | 1 | 1 | 0.436349766 |

| | | | | | |
|---|---|---|---|---|---|
| CD302 | CD302 molecule [Source:HGNC Symbol;Acc:HGNC:30843] | receptor | 0 | 0 | 0.436438041 |
| ELF4 | E74-like factor 4 (ets domain transcription factor) [Source:HGNC Symbol;Acc:HGNC:3319] | transcription | 0 | 1 | 0.436979249 |
| OAS2 | 2'-5'-oligoadenylate synthetase 2, 69/71kDa [Source:HGNC Symbol;Acc:HGNC:8087] | receptor | 1 | 1 | 0.437023238 |
| RNF31 | ring finger protein 31 [Source:HGNC Symbol;Acc:HGNC:16031] | accessory molecule | 0 | 1 | 0.437173085 |
| FGF20 | fibroblast growth factor 20 [Source:HGNC Symbol;Acc:HGNC:3677] | effector | 1 | 0 | 0.437396029 |
| C4BPB | complement component 4 binding protein, beta [Source:HGNC Symbol;Acc:HGNC:1328] | regulator | 1 | 1 | 0.437551373 |
| TRIM41 | tripartite motif containing 41 [Source:HGNC Symbol;Acc:HGNC:19013] | regulator | 0 | 0 | 0.437632837 |
| CXCR2 | chemokine (C-X-C motif) receptor 2 [Source:HGNC Symbol;Acc:HGNC:6027] | secondary receptor | 0 | 0 | 0.437738838 |
| CD27 | CD27 molecule [Source:HGNC Symbol;Acc:HGNC:11922] | secondary receptor | 0 | 1 | 0.437889959 |
| CD247 | CD247 molecule [Source:HGNC Symbol;Acc:HGNC:1677] | receptor | 1 | 0 | 0.437957339 |
| CCR2 | chemokine (C-C motif) receptor 2 [Source:HGNC Symbol;Acc:HGNC:1603] | secondary receptor | 1 | 0 | 0.438580741 |
| AZI2 | 5-azacytidine induced 2 [Source:HGNC Symbol;Acc:HGNC:24002] | adaptor | 0 | 1 | 0.438705071 |
| PLA2G6 | phospholipase A2, group VI (cytosolic, calcium-independent) [Source:HGNC Symbol;Acc:HGNC:9039] | signal transducer | 1 | 0 | 0.438761841 |
| IL17A | interleukin 17A [Source:HGNC Symbol;Acc:HGNC:5981] | effector | 0 | 1 | 0.438781879 |
| CEBPG | CCAAT/enhancer binding protein (C/EBP), gamma [Source:HGNC Symbol;Acc:HGNC:1837] | transcription | 1 | 0 | 0.438833425 |
| TREM1 | triggering receptor expressed on myeloid cells 1 [Source:HGNC Symbol;Acc:HGNC:17760] | receptor | 1 | 1 | 0.438852275 |
| NRG2 | neuregulin 2 [Source:HGNC Symbol;Acc:HGNC:7998] | UC | 1 | 0 | 0.439108801 |
| F11 | coagulation factor XI [Source:HGNC Symbol;Acc:HGNC:3529] | UC | 0 | 1 | 0.439476766 |
| HCK | HCK proto-oncogene, Src family tyrosine kinase [Source:HGNC Symbol;Acc:HGNC:4840] | signal transducer | 1 | 0 | 0.439534409 |
| FANCC | Fanconi anemia, complementation group C [Source:HGNC Symbol;Acc:HGNC:3584] | UC | 0 | 1 | 0.440200204 |
| IL22 | interleukin 22 [Source:HGNC Symbol;Acc:HGNC:14900] | effector | 0 | 1 | 0.440436551 |
| ITGAV | integrin, alpha V [Source:HGNC Symbol;Acc:HGNC:6150] | receptor | 0 | 1 | 0.440926378 |
| TRIM13 | tripartite motif containing 13 [Source:HGNC Symbol;Acc:HGNC:9976] | regulator | 1 | 1 | 0.44123433 |
| OTUB2 | OTU deubiquitinase, ubiquitin aldehyde binding 2 [Source:HGNC Symbol;Acc:HGNC:20351] | regulator | 0 | 1 | 0.441497497 |
| CSF2RB | colony stimulating factor 2 receptor, beta, low-affinity (granulocyte-macrophage) [Source:HGNC Symbol;Acc:HGNC:2436] | secondary receptor | 0 | 1 | 0.441909021 |
| IFNAR1 | interferon (alpha, beta and omega) receptor 1 [Source:HGNC Symbol;Acc:HGNC:5432] | secondary | 1 | 1 | 0.442009693 |

| | | | | | |
|---|---|---|---|---|---|
| | | receptor | | | |
| PLA2G4A | phospholipase A2, group IVA (cytosolic, calcium-dependent) [Source:HGNC Symbol;Acc:HGNC:9035] | effector | 0 | 1 | 0.442424273 |
| CCL17 | chemokine (C-C motif) ligand 17 [Source:HGNC Symbol;Acc:HGNC:10615] | effector | 0 | 1 | 0.442966177 |
| BIRC5 | baculoviral IAP repeat containing 5 [Source:HGNC Symbol;Acc:HGNC:593] | regulator | 0 | 1 | 0.443169807 |
| TRPM2 | transient receptor potential cation channel, subfamily M, member 2 [Source:HGNC Symbol;Acc:HGNC:12339] | UC | 0 | 1 | 0.443318318 |
| RNF216 | ring finger protein 216 [Source:HGNC Symbol;Acc:HGNC:21698] | regulator | 0 | 0 | 0.443447747 |
| FGF2 | fibroblast growth factor 2 (basic) [Source:HGNC Symbol;Acc:HGNC:3676] | effector | 1 | 0 | 0.44346062 |
| DEFB121 | defensin, beta 121 [Source:HGNC Symbol;Acc:HGNC:18101] | effector | 1 | 0 | 0.444462291 |
| TNF | tumor necrosis factor [Source:HGNC Symbol;Acc:HGNC:11892] | effector | 0 | 1 | 0.44513709 |
| MAPK13 | mitogen-activated protein kinase 13 [Source:HGNC Symbol;Acc:HGNC:6875] | signal transducer | 0 | 0 | 0.445435062 |
| ITGB3 | integrin, beta 3 (platelet glycoprotein IIIa, antigen CD61) [Source:HGNC Symbol;Acc:HGNC:6156] | receptor | 0 | 1 | 0.445592897 |
| TNFSF9 | tumor necrosis factor (ligand) superfamily, member 9 [Source:HGNC Symbol;Acc:HGNC:11939] | effector | 0 | 1 | 0.445853857 |
| CNOT4 | CCR4-NOT transcription complex, subunit 4 [Source:HGNC Symbol;Acc:HGNC:7880] | regulator | 0 | 1 | 0.446275166 |
| TRIB2 | tribbles pseudokinase 2 [Source:HGNC Symbol;Acc:HGNC:30809] | regulator | 0 | 1 | 0.446998047 |
| AKNA | AT-hook transcription factor [Source:HGNC Symbol;Acc:HGNC:24108] | transcription | 0 | 1 | 0.447252335 |
| TRIM36 | tripartite motif containing 36 [Source:HGNC Symbol;Acc:HGNC:16280] | regulator | 0 | 1 | 0.447593379 |
| AP3B1 | adaptor-related protein complex 3, beta 1 subunit [Source:HGNC Symbol;Acc:HGNC:566] | accessory molecule | 0 | 1 | 0.44788779 |
| IDO1 | indoleamine 2,3-dioxygenase 1 [Source:HGNC Symbol;Acc:HGNC:6059] | effector | 0 | 1 | 0.447942462 |
| TRIM14 | tripartite motif containing 14 [Source:HGNC Symbol;Acc:HGNC:16283] | regulator | 1 | 1 | 0.44794586 |
| MUL1 | mitochondrial E3 ubiquitin protein ligase 1 [Source:HGNC Symbol;Acc:HGNC:25762] | regulator | 0 | 1 | 0.447986376 |
| PTMA | prothymosin, alpha [Source:HGNC Symbol;Acc:HGNC:9623] | transcription | 0 | 1 | 0.448097851 |
| IRF9 | interferon regulatory factor 9 [Source:HGNC Symbol;Acc:HGNC:6131] | transcription | 1 | 1 | 0.448308501 |
| C1QBP | complement component 1, q subcomponent binding protein [Source:HGNC Symbol;Acc:HGNC:1243] | secondary receptor | 1 | 1 | 0.448313652 |
| WNT5A | wingless-type MMTV integration site family, member 5A [Source:HGNC Symbol;Acc:HGNC:12784] | effector | 1 | 0 | 0.448911025 |
| ISG20 | interferon stimulated exonuclease gene 20kDa [Source:HGNC Symbol;Acc:HGNC:6130] | effector | 1 | 1 | 0.449192505 |
| DAXX | death-domain associated protein [Source:HGNC Symbol;Acc:HGNC:2681] | transcription | 0 | 0 | 0.449258756 |
| LGALS9 | lectin, galactoside-binding, soluble, 9 [Source:HGNC Symbol;Acc:HGNC:6570] | receptor | 0 | 1 | 0.449395074 |
| APOH | apolipoprotein H (beta-2-glycoprotein I) [Source:HGNC Symbol;Acc:HGNC:616] | regulator | 0 | 1 | 0.449514612 |

| | | | | | |
|---|---|---|---|---|---|
| TBK1 | TANK-binding kinase 1 [Source:HGNC Symbol;Acc:HGNC:11584] | signal transducer | 1 | 1 | 0.449789936 |
| S100A12 | S100 calcium binding protein A12 [Source:HGNC Symbol;Acc:HGNC:10489] | effector | 1 | 0 | 0.449891738 |
| LTBP1 | latent transforming growth factor beta binding protein 1 [Source:HGNC Symbol;Acc:HGNC:6714] | regulator | 0 | 0 | 0.450008505 |
| IFIT5 | interferon-induced protein with tetratricopeptide repeats 5 [Source:HGNC Symbol;Acc:HGNC:13328] | receptor | 1 | 0 | 0.45038597 |
| USP21 | ubiquitin specific peptidase 21 [Source:HGNC Symbol;Acc:HGNC:12620] | regulator | 0 | 0 | 0.45048272 |
| TYRO3 | TYRO3 protein tyrosine kinase [Source:HGNC Symbol;Acc:HGNC:12446] | secondary receptor | 0 | 1 | 0.450569885 |
| IL9 | interleukin 9 [Source:HGNC Symbol;Acc:HGNC:6029] | effector | 0 | 1 | 0.450707366 |
| AICDA | activation-induced cytidine deaminase [Source:HGNC Symbol;Acc:HGNC:13203] | UC | 0 | 1 | 0.451214389 |
| IL6ST | interleukin 6 signal transducer [Source:HGNC Symbol;Acc:HGNC:6021] | secondary receptor | 0 | 1 | 0.451375671 |
| TRIM39 | tripartite motif containing 39 [Source:HGNC Symbol;Acc:HGNC:10065] | regulator | 0 | 0 | 0.451680736 |
| IL4R | interleukin 4 receptor [Source:HGNC Symbol;Acc:HGNC:6015] | secondary receptor | 0 | 1 | 0.451804221 |
| SNAP23 | synaptosomal-associated protein, 23kDa [Source:HGNC Symbol;Acc:HGNC:11131] | UC | 0 | 1 | 0.452224363 |
| TNFAIP8L2 | tumor necrosis factor, alpha-induced protein 8-like 2 [Source:HGNC Symbol;Acc:HGNC:26277] | regulator | 1 | 1 | 0.452324414 |
| C3AR1 | complement component 3a receptor 1 [Source:HGNC Symbol;Acc:HGNC:1319] | secondary receptor | 0 | 1 | 0.453272935 |
| CRP | C-reactive protein, pentraxin-related [Source:HGNC Symbol;Acc:HGNC:2367] | receptor | 0 | 1 | 0.453486542 |
| CAPZA1 | capping protein (actin filament) muscle Z-line, alpha 1 [Source:HGNC Symbol;Acc:HGNC:1488] | UC | 1 | 0 | 0.453772166 |
| MAP3K3 | mitogen-activated protein kinase kinase kinase 3 [Source:HGNC Symbol;Acc:HGNC:6855] | signal transducer | 0 | 1 | 0.453921726 |
| GAB2 | GRB2-associated binding protein 2 [Source:HGNC Symbol;Acc:HGNC:14458] | adaptor | 1 | 0 | 0.454061808 |
| PLA2G1B | phospholipase A2, group IB (pancreas) [Source:HGNC Symbol;Acc:HGNC:9030] | effector | 1 | 0 | 0.454086561 |
| DEFB115 | defensin, beta 115 [Source:HGNC Symbol;Acc:HGNC:18096] | effector | 1 | 0 | 0.455166659 |
| AIRE | autoimmune regulator [Source:HGNC Symbol;Acc:HGNC:360] | transcription | 0 | 1 | 0.455689505 |
| TRIM73 | tripartite motif containing 73 [Source:HGNC Symbol;Acc:HGNC:18162] | regulator | 0 | 0 | 0.455699966 |
| SMAD6 | SMAD family member 6 [Source:HGNC Symbol;Acc:HGNC:6772] | signal transducer | 0 | 1 | 0.455736942 |
| REL | v-rel avian reticuloendotheliosis viral oncogene homolog [Source:HGNC Symbol;Acc:HGNC:9954] | transcription | 1 | 1 | 0.456195367 |
| P2RY14 | purinergic receptor P2Y, G-protein coupled, 14 [Source:HGNC Symbol;Acc:HGNC:16442] | receptor | 0 | 1 | 0.456606155 |

| | | | | | |
|---|---|---|---|---|---|
| CD200 | CD200 molecule [Source:HGNC Symbol;Acc:HGNC:7203] | effector | 0 | 1 | 0.456901464 |
| PLK1 | polo-like kinase 1 [Source:HGNC Symbol;Acc:HGNC:9077] | regulator | 0 | 1 | 0.457044641 |
| ACHE | acetylcholinesterase (Yt blood group) [Source:HGNC Symbol;Acc:HGNC:108] | effector | 0 | 1 | 0.457232985 |
| ZC3H12A | zinc finger CCCH-type containing 12A [Source:HGNC Symbol;Acc:HGNC:26259] | effector | 0 | 1 | 0.457336236 |
| CARD9 | caspase recruitment domain family, member 9 [Source:HGNC Symbol;Acc:HGNC:16391] | adaptor | 1 | 1 | 0.457449151 |
| BST2 | bone marrow stromal cell antigen 2 [Source:HGNC Symbol;Acc:HGNC:1119] | effector | 1 | 1 | 0.457497848 |
| GOPC | golgi-associated PDZ and coiled-coil motif containing [Source:HGNC Symbol;Acc:HGNC:17643] | UC | 0 | 1 | 0.45797015 |
| LGR4 | leucine-rich repeat containing G protein-coupled receptor 4 [Source:HGNC Symbol;Acc:HGNC:13299] | secondary receptor | 1 | 0 | 0.457973505 |
| CORO2A | coronin, actin binding protein, 2A [Source:HGNC Symbol;Acc:HGNC:2255] | accessory molecule | 0 | 1 | 0.458471219 |
| CXCL1 | chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha) [Source:HGNC Symbol;Acc:HGNC:4602] | effector | 0 | 0 | 0.458569256 |
| IL1R1 | interleukin 1 receptor, type I [Source:HGNC Symbol;Acc:HGNC:5993] | secondary receptor | 0 | 1 | 0.458640656 |
| MASP1 | mannan-binding lectin serine peptidase 1 (C4/C2 activating component of Ra-reactive factor) [Source:HGNC Symbol;Acc:HGNC:6901] | adaptor | 1 | 1 | 0.458859824 |
| CNPY3 | canopy FGF signaling regulator 3 [Source:HGNC Symbol;Acc:HGNC:11968] | accessory molecule | 1 | 1 | 0.458946785 |
| DDX58 | DEAD (Asp-Glu-Ala-Asp) box polypeptide 58 [Source:HGNC Symbol;Acc:HGNC:19102] | receptor | 1 | 1 | 0.45899545 |
| TRPV2 | transient receptor potential cation channel, subfamily V, member 2 [Source:HGNC Symbol;Acc:HGNC:18082] | UC | 0 | 1 | 0.459691363 |
| PRKD1 | protein kinase D1 [Source:HGNC Symbol;Acc:HGNC:9407] | regulator | 1 | 0 | 0.460032173 |
| NPR1 | natriuretic peptide receptor 1 [Source:HGNC Symbol;Acc:HGNC:7943] | receptor | 0 | 0 | 0.460566741 |
| PYCARD | PYD and CARD domain containing [Source:HGNC Symbol;Acc:HGNC:16608] | adaptor | 1 | 1 | 0.460980548 |
| CD4 | CD4 molecule [Source:HGNC Symbol;Acc:HGNC:1678] | receptor | 1 | 0 | 0.460993934 |
| CD300LB | CD300 molecule-like family member b [Source:HGNC Symbol;Acc:HGNC:30811] | regulator | 1 | 0 | 0.461047425 |
| TP53 | tumor protein p53 [Source:HGNC Symbol;Acc:HGNC:11998] | transcription | 0 | 1 | 0.461141947 |
| MAPK11 | mitogen-activated protein kinase 11 [Source:HGNC Symbol;Acc:HGNC:6873] | signal transducer | 1 | 0 | 0.461232567 |
| EREG | epiregulin [Source:HGNC Symbol;Acc:HGNC:3443] | effector | 1 | 0 | 0.461250135 |
| CFD | complement factor D (adipsin) [Source:HGNC Symbol;Acc:HGNC:2771] | regulator | 1 | 0 | 0.46130453 |
| RB1 | retinoblastoma 1 [Source:HGNC Symbol;Acc:HGNC:9884] | regulator | 0 | 1 | 0.461824427 |

| | | | | | |
|---|---|---|---|---|---|
| CD180 | CD180 molecule [Source:HGNC Symbol;Acc:HGNC:6726] | receptor | 1 | 1 | 0.462006418 |
| TXN | thioredoxin [Source:HGNC Symbol;Acc:HGNC:12435] | regulator | 1 | 0 | 0.462370257 |
| PPARGC1A | peroxisome proliferator-activated receptor gamma, coactivator 1 alpha [Source:HGNC Symbol;Acc:HGNC:9237] | signal transducer | 0 | 1 | 0.462713768 |
| APCS | amyloid P component, serum [Source:HGNC Symbol;Acc:HGNC:584] | receptor | 1 | 1 | 0.463053924 |
| CCL8 | chemokine (C-C motif) ligand 8 [Source:HGNC Symbol;Acc:HGNC:10635] | effector | 0 | 0 | 0.463210999 |
| TRIM54 | tripartite motif containing 54 [Source:HGNC Symbol;Acc:HGNC:16008] | regulator | 0 | 0 | 0.463511258 |
| CRISP3 | cysteine-rich secretory protein 3 [Source:HGNC Symbol;Acc:HGNC:16904] | effector | 1 | 0 | 0.463529577 |
| ATG9A | autophagy related 9A [Source:HGNC Symbol;Acc:HGNC:22408] | regulator | 1 | 1 | 0.464081153 |
| CXCL10 | chemokine (C-X-C motif) ligand 10 [Source:HGNC Symbol;Acc:HGNC:10637] | effector | 0 | 1 | 0.465355824 |
| NPR3 | natriuretic peptide receptor 3 [Source:HGNC Symbol;Acc:HGNC:7945] | receptor | 0 | 0 | 0.465947572 |
| DEFB130 | defensin, beta 130 [Source:HGNC Symbol;Acc:HGNC:18107] | effector | 0 | 0 | 0.466771854 |
| IL12RB1 | interleukin 12 receptor, beta 1 [Source:HGNC Symbol;Acc:HGNC:5971] | secondary receptor | 1 | 0 | 0.466798697 |
| CLEC2B | C-type lectin domain family 2, member B [Source:HGNC Symbol;Acc:HGNC:2053] | receptor | 0 | 0 | 0.466820329 |
| PIK3CG | phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit gamma [Source:HGNC Symbol;Acc:HGNC:8978] | signal transducer | 1 | 0 | 0.467061395 |
| NLRP2 | NLR family, pyrin domain containing 2 [Source:HGNC Symbol;Acc:HGNC:22948] | receptor | 0 | 1 | 0.467127412 |
| KLK1 | kallikrein 1 [Source:HGNC Symbol;Acc:HGNC:6357] | signal transducer | 0 | 1 | 0.467450818 |
| E2F1 | E2F transcription factor 1 [Source:HGNC Symbol;Acc:HGNC:3113] | regulator | 0 | 1 | 0.467461908 |
| PTPN2 | protein tyrosine phosphatase, non-receptor type 2 [Source:HGNC Symbol;Acc:HGNC:9650] | regulator | 1 | 1 | 0.468128248 |
| NUP214 | nucleoporin 214kDa [Source:HGNC Symbol;Acc:HGNC:8064] | accessory molecule | 0 | 1 | 0.468581395 |
| NRROS | negative regulator of reactive oxygen species [Source:HGNC Symbol;Acc:HGNC:24613] | regulator | 1 | 0 | 0.468612631 |
| ERN1 | endoplasmic reticulum to nucleus signaling 1 [Source:HGNC Symbol;Acc:HGNC:3449] | UC | 0 | 1 | 0.468772581 |
| NFKBIZ | nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, zeta [Source:HGNC Symbol;Acc:HGNC:29805] | signal transducer | 0 | 1 | 0.468777176 |
| EYA4 | EYA transcriptional coactivator and phosphatase 4 [Source:HGNC Symbol;Acc:HGNC:3522] | regulator | 0 | 0 | 0.469374884 |
| RNF125 | ring finger protein 125, E3 ubiquitin protein ligase [Source:HGNC Symbol;Acc:HGNC:21150] | regulator | 1 | 1 | 0.469525792 |
| DUSP10 | dual specificity phosphatase 10 [Source:HGNC Symbol;Acc:HGNC:3065] | regulator | 0 | 1 | 0.469586459 |
| IL5 | interleukin 5 [Source:HGNC Symbol;Acc:HGNC:6016] | effector | 0 | 1 | 0.469654693 |
| ING4 | inhibitor of growth family, member 4 [Source:HGNC Symbol;Acc:HGNC:19423] | signal | 0 | 1 | 0.469778853 |

| | | | | | |
|---|---|---|---|---|---|
| PMAIP1 | phorbol-12-myristate-13-acetate-induced protein 1 [Source:HGNC Symbol;Acc:HGNC:9108] | transducer signal transducer | 0 | 1 | 0.469866189 |
| PTGES | prostaglandin E synthase [Source:HGNC Symbol;Acc:HGNC:9599] | effector | 0 | 1 | 0.47015177 |
| CEBPE | CCAAT/enhancer binding protein (C/EBP), epsilon [Source:HGNC Symbol;Acc:HGNC:1836] | transcription | 0 | 1 | 0.470258226 |
| IL31 | interleukin 31 [Source:HGNC Symbol;Acc:HGNC:19372] | effector | 0 | 1 | 0.470789689 |
| IFNL2 | interferon, lambda 2 [Source:HGNC Symbol;Acc:HGNC:18364] | effector | 0 | 0 | 0.470868645 |
| MYD88 | myeloid differentiation primary response 88 [Source:HGNC Symbol;Acc:HGNC:7562] | adaptor | 1 | 1 | 0.47116945 |
| ITGA3 | integrin, alpha 3 (antigen CD49C, alpha 3 subunit of VLA-3 receptor) [Source:HGNC Symbol;Acc:HGNC:6139] | receptor | 0 | 1 | 0.471261576 |
| TRADD | TNFRSF1A-associated via death domain [Source:HGNC Symbol;Acc:HGNC:12030] | signal transducer | 0 | 1 | 0.471416436 |
| POLR3K | polymerase (RNA) III (DNA directed) polypeptide K, 12.3 kDa [Source:HGNC Symbol;Acc:HGNC:14121] | receptor | 1 | 0 | 0.471821953 |
| IFIT3 | interferon-induced protein with tetratricopeptide repeats 3 [Source:HGNC Symbol;Acc:HGNC:5411] | adaptor | 1 | 1 | 0.47270066 |
| BLNK | B-cell linker [Source:HGNC Symbol;Acc:HGNC:14211] | signal transducer | 0 | 0 | 0.4727971 |
| SPINK5 | serine peptidase inhibitor, Kazal type 5 [Source:HGNC Symbol;Acc:HGNC:15464] | adaptor | 0 | 1 | 0.473211496 |
| C1R | complement component 1, r subcomponent [Source:HGNC Symbol;Acc:HGNC:1246] | adaptor | 1 | 1 | 0.47332315 |
| NLRP5 | NLR family, pyrin domain containing 5 [Source:HGNC Symbol;Acc:HGNC:21269] | receptor | 0 | 1 | 0.473560955 |
| IL1B | interleukin 1, beta [Source:HGNC Symbol;Acc:HGNC:5992] | effector | 0 | 1 | 0.474093639 |
| IGF1 | insulin-like growth factor 1 (somatomedin C) [Source:HGNC Symbol;Acc:HGNC:5464] | effector | 0 | 1 | 0.474272238 |
| WIPF3 | WAS/WASL interacting protein family, member 3 [Source:HGNC Symbol;Acc:HGNC:22004] | UC | 1 | 0 | 0.474513927 |
| PLAUR | plasminogen activator, urokinase receptor [Source:HGNC Symbol;Acc:HGNC:9053] | receptor | 0 | 1 | 0.474834074 |
| ACE2 | angiotensin I converting enzyme 2 [Source:HGNC Symbol;Acc:HGNC:13557] | regulator | 0 | 1 | 0.475816545 |
| SPHK1 | sphingosine kinase 1 [Source:HGNC Symbol;Acc:HGNC:11240] | signal transducer | 0 | 1 | 0.476054633 |
| SLC15A4 | solute carrier family 15 (oligopeptide transporter), member 4 [Source:HGNC Symbol;Acc:HGNC:23090] | accessory molecule | 0 | 1 | 0.47635246 |
| MX2 | MX dynamin-like GTPase 2 [Source:HGNC Symbol;Acc:HGNC:7533] | effector | 1 | 1 | 0.476642421 |
| SSC5D | scavenger receptor cysteine rich family, 5 domains [Source:HGNC Symbol;Acc:HGNC:26641] | receptor | 1 | 0 | 0.477198437 |
| MFN1 | mitofusin 1 [Source:HGNC Symbol;Acc:HGNC:18262] | accessory molecule | 0 | 1 | 0.477486132 |

| | | | | | |
|---|---|---|---|---|---|
| FKBP5 | FK506 binding protein 5 [Source:HGNC Symbol;Acc:HGNC:3721] | regulator | 0 | 0 | 0.477509921 |
| RHBDF2 | rhomboid 5 homolog 2 (Drosophila) [Source:HGNC Symbol;Acc:HGNC:20788] | accessory molecule | 0 | 1 | 0.477583043 |
| FCGR2A | Fc fragment of IgG, low affinity IIa, receptor (CD32) [Source:HGNC Symbol;Acc:HGNC:3616] | receptor | 1 | 1 | 0.477756074 |
| MAP4K2 | mitogen-activated protein kinase kinase kinase kinase 2 [Source:HGNC Symbol;Acc:HGNC:6864] | signal transducer | 1 | 0 | 0.478010228 |
| ELMOD2 | ELMO/CED-12 domain containing 2 [Source:HGNC Symbol;Acc:HGNC:28111] | UC | 0 | 1 | 0.478747686 |
| RIPK3 | receptor-interacting serine-threonine kinase 3 [Source:HGNC Symbol;Acc:HGNC:10021] | signal transducer | 1 | 1 | 0.478764265 |
| CD46 | CD46 molecule, complement regulatory protein [Source:HGNC Symbol;Acc:HGNC:6953] | receptor | 1 | 1 | 0.478849091 |
| ARHGEF12 | Rho guanine nucleotide exchange factor (GEF) 12 [Source:HGNC Symbol;Acc:HGNC:14193] | signal transducer | 0 | 0 | 0.479239134 |
| TNK1 | tyrosine kinase, non-receptor, 1 [Source:HGNC Symbol;Acc:HGNC:11940] | signal transducer | 1 | 1 | 0.479397374 |
| DDX21 | DEAD (Asp-Glu-Ala-Asp) box helicase 21 [Source:HGNC Symbol;Acc:HGNC:2744] | adaptor | 0 | 1 | 0.479528898 |
| EPRS | glutamyl-prolyl-tRNA synthetase [Source:HGNC Symbol;Acc:HGNC:3418] | regulator | 1 | 0 | 0.479586233 |
| AHR | aryl hydrocarbon receptor [Source:HGNC Symbol;Acc:HGNC:348] | UC | 0 | 1 | 0.479591809 |
| TRIM25 | tripartite motif containing 25 [Source:HGNC Symbol;Acc:HGNC:12932] | regulator | 1 | 1 | 0.479680907 |
| SIRT2 | sirtuin 2 [Source:HGNC Symbol;Acc:HGNC:10886] | UC | 1 | 0 | 0.479959959 |
| POLR2H | polymerase (RNA) II (DNA directed) polypeptide H [Source:HGNC Symbol;Acc:HGNC:9195] | receptor | 1 | 0 | 0.480201472 |
| NLRP11 | NLR family, pyrin domain containing 11 [Source:HGNC Symbol;Acc:HGNC:22945] | receptor | 0 | 1 | 0.480478638 |
| CALCOCO2 | calcium binding and coiled-coil domain 2 [Source:HGNC Symbol;Acc:HGNC:29912] | receptor | 1 | 1 | 0.48059546 |
| CCL1 | chemokine (C-C motif) ligand 1 [Source:HGNC Symbol;Acc:HGNC:10609] | effector | 0 | 1 | 0.480691021 |
| HSPA14 | heat shock 70kDa protein 14 [Source:HGNC Symbol;Acc:HGNC:29526] | UC | 0 | 1 | 0.480800186 |
| FADD | Fas (TNFRSF6)-associated via death domain [Source:HGNC Symbol;Acc:HGNC:3573] | signal transducer | 1 | 1 | 0.480939949 |
| CTSS | cathepsin S [Source:HGNC Symbol;Acc:HGNC:2545] | accessory molecule | 1 | 1 | 0.480960046 |
| LAT2 | linker for activation of T cells family, member 2 [Source:HGNC Symbol;Acc:HGNC:12749] | signal transducer | 1 | 0 | 0.481058262 |
| EIF2AK1 | eukaryotic translation initiation factor 2-alpha kinase 1 [Source:HGNC Symbol;Acc:HGNC:24921] | receptor | 0 | 0 | 0.481635544 |
| CXCL12 | chemokine (C-X-C motif) ligand 12 [Source:HGNC Symbol;Acc:HGNC:10672] | effector | 0 | 1 | 0.481870253 |
| CD55 | CD55 molecule, decay accelerating factor for complement (Cromer blood group) [Source:HGNC | receptor | 1 | 0 | 0.482102587 |

| | | | | | |
|---|---|---|---|---|---|
| | Symbol;Acc:HGNC:2665] | | | | |
| TXNIP | thioredoxin interacting protein [Source:HGNC Symbol;Acc:HGNC:16952] | signal transducer | 1 | 1 | 0.482165345 |
| NLRP13 | NLR family, pyrin domain containing 13 [Source:HGNC Symbol;Acc:HGNC:22937] | receptor | 0 | 1 | 0.482232806 |
| CBL | Cbl proto-oncogene, E3 ubiquitin protein ligase [Source:HGNC Symbol;Acc:HGNC:1541] | regulator | 0 | 1 | 0.482464218 |
| SERPING1 | serpin peptidase inhibitor, clade G (C1 inhibitor), member 1 [Source:HGNC Symbol;Acc:HGNC:1228] | regulator | 1 | 1 | 0.482487309 |
| HCST | hematopoietic cell signal transducer [Source:HGNC Symbol;Acc:HGNC:16977] | adaptor | 0 | 0 | 0.482663482 |
| DEFB136 | defensin, beta 136 [Source:HGNC Symbol;Acc:HGNC:34433] | effector | 0 | 0 | 0.482792721 |
| MST1R | macrophage stimulating 1 receptor (c-met-related tyrosine kinase) [Source:HGNC Symbol;Acc:HGNC:7381] | receptor | 1 | 1 | 0.48320968 |
| MARCO | macrophage receptor with collagenous structure [Source:HGNC Symbol;Acc:HGNC:6895] | receptor | 1 | 1 | 0.483251915 |
| IL23R | interleukin 23 receptor [Source:HGNC Symbol;Acc:HGNC:19100] | secondary receptor | 1 | 0 | 0.483474564 |
| CHID1 | chitinase domain containing 1 [Source:HGNC Symbol;Acc:HGNC:28474] | receptor | 1 | 0 | 0.483512338 |
| PRDX1 | peroxiredoxin 1 [Source:HGNC Symbol;Acc:HGNC:9352] | UC | 1 | 0 | 0.48377304 |
| COCH | cochlin [Source:HGNC Symbol;Acc:HGNC:2180] | effector | 0 | 0 | 0.483890937 |
| C1RL | complement component 1, r subcomponent-like [Source:HGNC Symbol;Acc:HGNC:21265] | signal transducer | 1 | 0 | 0.484726472 |
| CLU | clusterin [Source:HGNC Symbol;Acc:HGNC:2095] | accessory molecule | 1 | 0 | 0.485119568 |
| ITGAM | integrin, alpha M (complement component 3 receptor 3 subunit) [Source:HGNC Symbol;Acc:HGNC:6149] | receptor | 1 | 1 | 0.485144951 |
| SIRPA | signal-regulatory protein alpha [Source:HGNC Symbol;Acc:HGNC:9662] | effector | 0 | 1 | 0.485239104 |
| ANXA4 | annexin A4 [Source:HGNC Symbol;Acc:HGNC:542] | regulator | 0 | 1 | 0.485365816 |
| CLEC18B | C-type lectin domain family 18, member B [Source:HGNC Symbol;Acc:HGNC:33849] | receptor | 0 | 0 | 0.485699294 |
| FGF3 | fibroblast growth factor 3 [Source:HGNC Symbol;Acc:HGNC:3681] | effector | 1 | 0 | 0.485785783 |
| CLEC4G | C-type lectin domain family 4, member G [Source:HGNC Symbol;Acc:HGNC:24591] | receptor | 0 | 0 | 0.485811406 |
| NLRP7 | NLR family, pyrin domain containing 7 [Source:HGNC Symbol;Acc:HGNC:22947] | receptor | 0 | 1 | 0.486098015 |
| CTSK | cathepsin K [Source:HGNC Symbol;Acc:HGNC:2536] | accessory molecule | 1 | 1 | 0.486448819 |
| CR1 | complement component (3b/4b) receptor 1 (Knops blood group) [Source:HGNC Symbol;Acc:HGNC:2334] | secondary receptor | 1 | 0 | 0.486837271 |
| VDR | vitamin D (1,25- dihydroxyvitamin D3) receptor [Source:HGNC Symbol;Acc:HGNC:12679] | secondary | 0 | 1 | 0.486869081 |

| | | | | | |
|---|---|---|---|---|---|
| | | receptor | | | |
| ASGR1 | asialoglycoprotein receptor 1 [Source:HGNC Symbol;Acc:HGNC:742] | receptor | 0 | 0 | 0.486943449 |
| TDGF1 | teratocarcinoma-derived growth factor 1 [Source:HGNC Symbol;Acc:HGNC:11701] | UC | 1 | 0 | 0.487315287 |
| HAMP | hepcidin antimicrobial peptide [Source:HGNC Symbol;Acc:HGNC:15598] | effector | 0 | 1 | 0.487354407 |
| CLEC4A | C-type lectin domain family 4, member A [Source:HGNC Symbol;Acc:HGNC:13257] | receptor | 1 | 0 | 0.487436526 |
| PARD3 | par-3 family cell polarity regulator [Source:HGNC Symbol;Acc:HGNC:16051] | regulator | 0 | 1 | 0.487563649 |
| BLK | BLK proto-oncogene, Src family tyrosine kinase [Source:HGNC Symbol;Acc:HGNC:1057] | signal transducer | 1 | 0 | 0.487819914 |
| TLR2 | toll-like receptor 2 [Source:HGNC Symbol;Acc:HGNC:11848] | receptor | 1 | 1 | 0.487955782 |
| IL19 | interleukin 19 [Source:HGNC Symbol;Acc:HGNC:5990] | effector | 0 | 1 | 0.487994498 |
| WIPF2 | WAS/WASL interacting protein family, member 2 [Source:HGNC Symbol;Acc:HGNC:30923] | UC | 1 | 0 | 0.488350785 |
| CD19 | CD19 molecule [Source:HGNC Symbol;Acc:HGNC:1633] | receptor | 1 | 0 | 0.489178637 |
| FREM1 | FRAS1 related extracellular matrix 1 [Source:HGNC Symbol;Acc:HGNC:23399] | UC | 0 | 1 | 0.489557133 |
| BTN3A3 | butyrophilin, subfamily 3, member A3 [Source:HGNC Symbol;Acc:HGNC:1140] | receptor | 0 | 1 | 0.489756189 |
| RAG1 | recombination activating gene 1 [Source:HGNC Symbol;Acc:HGNC:9831] | UC | 0 | 1 | 0.490268645 |
| GAS6 | growth arrest-specific 6 [Source:HGNC Symbol;Acc:HGNC:4168] | adaptor | 0 | 1 | 0.490371512 |
| BCL10 | B-cell CLL/lymphoma 10 [Source:HGNC Symbol;Acc:HGNC:989] | signal transducer | 1 | 1 | 0.490412767 |
| CTSH | cathepsin H [Source:HGNC Symbol;Acc:HGNC:2535] | accessory molecule | 0 | 1 | 0.490720996 |
| SIGLEC1 | sialic acid binding Ig-like lectin 1, sialoadhesin [Source:HGNC Symbol;Acc:HGNC:11127] | receptor | 0 | 1 | 0.490879605 |
| CDKN1B | cyclin-dependent kinase inhibitor 1B (p27, Kip1) [Source:HGNC Symbol;Acc:HGNC:1785] | signal transducer | 1 | 0 | 0.491030427 |
| NLRC4 | NLR family, CARD domain containing 4 [Source:HGNC Symbol;Acc:HGNC:16412] | receptor | 1 | 1 | 0.491062255 |
| TOLLIP | toll interacting protein [Source:HGNC Symbol;Acc:HGNC:16476] | regulator | 1 | 1 | 0.492638551 |
| NRG4 | neuregulin 4 [Source:HGNC Symbol;Acc:HGNC:29862] | UC | 1 | 0 | 0.493233159 |
| TRIM32 | tripartite motif containing 32 [Source:HGNC Symbol;Acc:HGNC:16380] | regulator | 1 | 1 | 0.493634631 |
| CLEC7A | C-type lectin domain family 7, member A [Source:HGNC Symbol;Acc:HGNC:14558] | receptor | 1 | 1 | 0.494027068 |
| FSTL1 | follistatin-like 1 [Source:HGNC Symbol;Acc:HGNC:3972] | effector | 0 | 1 | 0.495589462 |
| FCN1 | ficolin (collagen/fibrinogen domain containing) 1 [Source:HGNC Symbol;Acc:HGNC:3623] | receptor | 1 | 1 | 0.495853598 |
| TNRC6B | trinucleotide repeat containing 6B [Source:HGNC Symbol;Acc:HGNC:29190] | regulator | 1 | 0 | 0.495979254 |
| FCER1A | Fc fragment of IgE, high affinity I, receptor for; alpha polypeptide [Source:HGNC | receptor | 1 | 0 | 0.496186021 |

| | | | | | |
|---|---|---|---|---|---|
| | Symbol;Acc:HGNC:3609] | | | | |
| HERC5 | HECT and RLD domain containing E3 ubiquitin protein ligase 5 [Source:HGNC Symbol;Acc:HGNC:24368] | regulator | 1 | 1 | 0.496283924 |
| ATF3 | activating transcription factor 3 [Source:HGNC Symbol;Acc:HGNC:785] | transcription | 0 | 1 | 0.49648481 |
| MAP2K3 | mitogen-activated protein kinase kinase 3 [Source:HGNC Symbol;Acc:HGNC:6843] | signal transducer | 1 | 1 | 0.496703783 |
| IRF7 | interferon regulatory factor 7 [Source:HGNC Symbol;Acc:HGNC:6122] | transcription | 1 | 1 | 0.496723126 |
| EGF | epidermal growth factor [Source:HGNC Symbol;Acc:HGNC:3229] | effector | 1 | 0 | 0.49674585 |
| CFP | complement factor properdin [Source:HGNC Symbol;Acc:HGNC:8864] | regulator | 1 | 1 | 0.496997243 |
| ADAM15 | ADAM metallopeptidase domain 15 [Source:HGNC Symbol;Acc:HGNC:193] | regulator | 1 | 0 | 0.497318425 |
| DUSP6 | dual specificity phosphatase 6 [Source:HGNC Symbol;Acc:HGNC:3072] | regulator | 1 | 0 | 0.497594773 |
| IL4 | interleukin 4 [Source:HGNC Symbol;Acc:HGNC:6014] | effector | 1 | 1 | 0.49796946 |
| AGER | advanced glycosylation end product-specific receptor [Source:HGNC Symbol;Acc:HGNC:320] | receptor | 1 | 1 | 0.498486644 |
| CASP4 | caspase 4, apoptosis-related cysteine peptidase [Source:HGNC Symbol;Acc:HGNC:1505] | signal transducer | 0 | 0 | 0.498502694 |
| PROCR | protein C receptor, endothelial [Source:HGNC Symbol;Acc:HGNC:9452] | UC | 0 | 1 | 0.49854807 |
| SHC1 | SHC (Src homology 2 domain containing) transforming protein 1 [Source:HGNC Symbol;Acc:HGNC:10840] | adaptor | 1 | 0 | 0.498650189 |
| TRIM38 | tripartite motif containing 38 [Source:HGNC Symbol;Acc:HGNC:10059] | regulator | 1 | 1 | 0.499171507 |
| TRIM52 | tripartite motif containing 52 [Source:HGNC Symbol;Acc:HGNC:19024] | regulator | 0 | 0 | 0.499472752 |
| TBKBP1 | TBK1 binding protein 1 [Source:HGNC Symbol;Acc:HGNC:30140] | adaptor | 1 | 1 | 0.499527338 |
| COPS8 | COP9 signalosome subunit 8 [Source:HGNC Symbol;Acc:HGNC:24335] | signal transducer | 0 | 1 | 0.499543499 |
| PIAS3 | protein inhibitor of activated STAT, 3 [Source:HGNC Symbol;Acc:HGNC:16861] | regulator | 0 | 1 | 0.499836162 |
| SIRPB1 | signal-regulatory protein beta 1 [Source:HGNC Symbol;Acc:HGNC:15928] | secondary receptor | 1 | 0 | 0.499844994 |
| MAP3K12 | mitogen-activated protein kinase kinase kinase 12 [Source:HGNC Symbol;Acc:HGNC:6851] | signal transducer | 0 | 1 | 0.500152149 |
| CLEC4C | C-type lectin domain family 4, member C [Source:HGNC Symbol;Acc:HGNC:13258] | receptor | 1 | 1 | 0.500206006 |
| ADCY6 | adenylate cyclase 6 [Source:HGNC Symbol;Acc:HGNC:237] | UC | 1 | 0 | 0.500346315 |
| POLR3G | polymerase (RNA) III (DNA directed) polypeptide G (32kD) [Source:HGNC Symbol;Acc:HGNC:30075] | receptor | 1 | 0 | 0.500647268 |
| ACE | angiotensin I converting enzyme [Source:HGNC Symbol;Acc:HGNC:2707] | regulator | 0 | 0 | 0.500786297 |
| DEFB131 | defensin, beta 131 [Source:HGNC Symbol;Acc:HGNC:18108] | effector | 1 | 0 | 0.500908219 |

| | | | | | |
|---|---|---|---|---|---|
| PLSCR1 | phospholipid scramblase 1 [Source:HGNC Symbol;Acc:HGNC:9092] | accessory molecule | 0 | 1 | 0.500927807 |
| NOXA1 | NADPH oxidase activator 1 [Source:HGNC Symbol;Acc:HGNC:10668] | effector | 0 | 1 | 0.501040572 |
| PLTP | phospholipid transfer protein [Source:HGNC Symbol;Acc:HGNC:9093] | UC | 0 | 1 | 0.501059407 |
| PF4 | platelet factor 4 [Source:HGNC Symbol;Acc:HGNC:8861] | effector | 0 | 0 | 0.501209483 |
| PIK3R2 | phosphoinositide-3-kinase, regulatory subunit 2 (beta) [Source:HGNC Symbol;Acc:HGNC:8980] | signal transducer | 1 | 0 | 0.50122076 |
| COX5B | cytochrome c oxidase subunit Vb [Source:HGNC Symbol;Acc:HGNC:2269] | UC | 0 | 1 | 0.501258502 |
| WASF2 | WAS protein family, member 2 [Source:HGNC Symbol;Acc:HGNC:12733] | UC | 1 | 0 | 0.502205488 |
| AXL | AXL receptor tyrosine kinase [Source:HGNC Symbol;Acc:HGNC:905] | effector | 1 | 1 | 0.502599306 |
| XIAP | X-linked inhibitor of apoptosis, E3 ubiquitin protein ligase [Source:HGNC Symbol;Acc:HGNC:592] | regulator | 0 | 1 | 0.502621654 |
| LGALS8 | lectin, galactoside-binding, soluble, 8 [Source:HGNC Symbol;Acc:HGNC:6569] | receptor | 0 | 1 | 0.502645001 |
| IFNW1 | interferon, omega 1 [Source:HGNC Symbol;Acc:HGNC:5448] | effector | 1 | 0 | 0.502760056 |
| MRE11A | MRE11 meiotic recombination 11 homolog A (S. cerevisiae) [Source:HGNC Symbol;Acc:HGNC:7230] | receptor | 1 | 0 | 0.502946272 |
| CLEC14A | C-type lectin domain family 14, member A [Source:HGNC Symbol;Acc:HGNC:19832] | receptor | 0 | 0 | 0.502983413 |
| VPS45 | vacuolar protein sorting 45 homolog (S. cerevisiae) [Source:HGNC Symbol;Acc:HGNC:14579] | accessory molecule | 0 | 1 | 0.503044314 |
| ABI1 | abl-interactor 1 [Source:HGNC Symbol;Acc:HGNC:11320] | UC | 1 | 0 | 0.503285132 |
| PTAFR | platelet-activating factor receptor [Source:HGNC Symbol;Acc:HGNC:9582] | secondary receptor | 1 | 1 | 0.503286715 |
| TRIM51 | tripartite motif-containing 51 [Source:HGNC Symbol;Acc:HGNC:19023] | regulator | 0 | 0 | 0.503636949 |
| CTNNAL1 | catenin (cadherin-associated protein), alpha-like 1 [Source:HGNC Symbol;Acc:HGNC:2512] | UC | 0 | 1 | 0.503792353 |
| CD63 | CD63 molecule [Source:HGNC Symbol;Acc:HGNC:1692] | receptor | 0 | 1 | 0.504267781 |
| SIGLEC15 | sialic acid binding Ig-like lectin 15 [Source:HGNC Symbol;Acc:HGNC:27596] | receptor | 1 | 1 | 0.504782323 |
| TNFRSF9 | tumor necrosis factor receptor superfamily, member 9 [Source:HGNC Symbol;Acc:HGNC:11924] | secondary receptor | 0 | 1 | 0.505241626 |
| TRIM7 | tripartite motif containing 7 [Source:HGNC Symbol;Acc:HGNC:16278] | regulator | 0 | 1 | 0.505389748 |
| TRIM65 | tripartite motif containing 65 [Source:HGNC Symbol;Acc:HGNC:27316] | regulator | 0 | 1 | 0.506020124 |
| CD69 | CD69 molecule [Source:HGNC Symbol;Acc:HGNC:1694] | signal transducer | 0 | 0 | 0.506139947 |
| ELF1 | E74-like factor 1 (ets domain transcription factor) [Source:HGNC Symbol;Acc:HGNC:3316] | transcription | 0 | 1 | 0.506201621 |
| MAPK7 | mitogen-activated protein kinase 7 [Source:HGNC Symbol;Acc:HGNC:6880] | signal transducer | 1 | 0 | 0.506274417 |

| | | | | | |
|---|---|---|---|---|---|
| IL1RL2 | interleukin 1 receptor-like 2 [Source:HGNC Symbol;Acc:HGNC:5999] | secondary receptor | 1 | 1 | 0.506281092 |
| IL18R1 | interleukin 18 receptor 1 [Source:HGNC Symbol;Acc:HGNC:5988] | secondary receptor | 0 | 1 | 0.506545478 |
| CLEC2A | C-type lectin domain family 2, member A [Source:HGNC Symbol;Acc:HGNC:24191] | receptor | 1 | 0 | 0.50687911 |
| MAVS | mitochondrial antiviral signaling protein [Source:HGNC Symbol;Acc:HGNC:29233] | adaptor | 1 | 1 | 0.50688036 |
| C2 | complement component 2 [Source:HGNC Symbol;Acc:HGNC:1248] | signal transducer | 1 | 1 | 0.507349471 |
| FGF10 | fibroblast growth factor 10 [Source:HGNC Symbol;Acc:HGNC:3666] | effector | 1 | 0 | 0.507407394 |
| PRKAR2A | protein kinase, cAMP-dependent, regulatory, type II, alpha [Source:HGNC Symbol;Acc:HGNC:9391] | regulator | 1 | 0 | 0.50909922 |
| C5AR1 | complement component 5a receptor 1 [Source:HGNC Symbol;Acc:HGNC:1338] | secondary receptor | 0 | 1 | 0.509254292 |
| IFNK | interferon, kappa [Source:HGNC Symbol;Acc:HGNC:21714] | effector | 1 | 0 | 0.509327111 |
| NFATC4 | nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 4 [Source:HGNC Symbol;Acc:HGNC:7778] | transcription | 0 | 1 | 0.509974204 |
| NLRP9 | NLR family, pyrin domain containing 9 [Source:HGNC Symbol;Acc:HGNC:22941] | receptor | 0 | 1 | 0.509990518 |
| ELANE | elastase, neutrophil expressed [Source:HGNC Symbol;Acc:HGNC:3309] | regulator | 0 | 1 | 0.509994459 |
| SIGLEC5 | sialic acid binding Ig-like lectin 5 [Source:HGNC Symbol;Acc:HGNC:10874] | receptor | 0 | 1 | 0.510004283 |
| IFITM2 | interferon induced transmembrane protein 2 [Source:HGNC Symbol;Acc:HGNC:5413] | effector | 1 | 1 | 0.510704573 |
| SARM1 | sterile alpha and TIR motif containing 1 [Source:HGNC Symbol;Acc:HGNC:17074] | adaptor | 1 | 1 | 0.510904238 |
| NLRC5 | NLR family, CARD domain containing 5 [Source:HGNC Symbol;Acc:HGNC:29933] | receptor | 1 | 1 | 0.510907451 |
| PTPRC | protein tyrosine phosphatase, receptor type, C [Source:HGNC Symbol;Acc:HGNC:9666] | regulator | 0 | 1 | 0.511365161 |
| REST | RE1-silencing transcription factor [Source:HGNC Symbol;Acc:HGNC:9966] | regulator | 0 | 1 | 0.511540803 |
| CFH | complement factor H [Source:HGNC Symbol;Acc:HGNC:4883] | regulator | 1 | 1 | 0.511762894 |
| AIM2 | absent in melanoma 2 [Source:HGNC Symbol;Acc:HGNC:357] | receptor | 1 | 1 | 0.512378662 |
| APOA1 | apolipoprotein A-I [Source:HGNC Symbol;Acc:HGNC:600] | regulator | 0 | 1 | 0.512410604 |
| LPCAT2 | lysophosphatidylcholine acyltransferase 2 [Source:HGNC Symbol;Acc:HGNC:26032] | signal transducer | 0 | 1 | 0.512688862 |
| PALM3 | paralemmin 3 [Source:HGNC Symbol;Acc:HGNC:33274] | adaptor | 0 | 0 | 0.512899669 |
| SERPINE1 | serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1 [Source:HGNC Symbol;Acc:HGNC:8583] | regulator | 0 | 1 | 0.512945105 |
| SLAMF6 | SLAM family member 6 [Source:HGNC Symbol;Acc:HGNC:21392] | receptor | 0 | 1 | 0.513901538 |
| CD37 | CD37 molecule [Source:HGNC Symbol;Acc:HGNC:1666] | accessory | 0 | 1 | 0.513937991 |

| | | | | | |
|---|---|---|---|---|---|
| | | molecule | | | |
| VEGFA | vascular endothelial growth factor A [Source:HGNC Symbol;Acc:HGNC:12680] | effector | 0 | 1 | 0.514002498 |
| NFATC2 | nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 2 [Source:HGNC Symbol;Acc:HGNC:7776] | transcription | 1 | 1 | 0.514007983 |
| F2RL1 | coagulation factor II (thrombin) receptor-like 1 [Source:HGNC Symbol;Acc:HGNC:3538] | secondary receptor | 1 | 1 | 0.514557333 |
| PYHIN1 | pyrin and HIN domain family, member 1 [Source:HGNC Symbol;Acc:HGNC:28894] | receptor | 0 | 0 | 0.514762869 |
| TANK | TRAF family member-associated NFKB activator [Source:HGNC Symbol;Acc:HGNC:11562] | signal transducer | 1 | 1 | 0.514953825 |
| TRIML1 | tripartite motif family-like 1 [Source:HGNC Symbol;Acc:HGNC:26698] | regulator | 0 | 0 | 0.515198744 |
| LILRA4 | leukocyte immunoglobulin-like receptor, subfamily A (with TM domain), member 4 [Source:HGNC Symbol;Acc:HGNC:15503] | receptor | 0 | 1 | 0.515310706 |
| GP2 | glycoprotein 2 (zymogen granule membrane) [Source:HGNC Symbol;Acc:HGNC:4441] | receptor | 0 | 1 | 0.515338392 |
| WDR62 | WD repeat domain 62 [Source:HGNC Symbol;Acc:HGNC:24502] | signal transducer | 0 | 1 | 0.515476345 |
| NLRP10 | NLR family, pyrin domain containing 10 [Source:HGNC Symbol;Acc:HGNC:21464] | receptor | 1 | 1 | 0.515507828 |
| C1QA | complement component 1, q subcomponent, A chain [Source:HGNC Symbol;Acc:HGNC:1241] | receptor | 1 | 1 | 0.515639434 |
| IL17C | interleukin 17C [Source:HGNC Symbol;Acc:HGNC:5983] | effector | 0 | 1 | 0.515959518 |
| CASP12 | caspase 12 (gene/pseudogene) [Source:HGNC Symbol;Acc:HGNC:19004] | signal transducer | 0 | 1 | 0.516083934 |
| MIF | macrophage migration inhibitory factor (glycosylation-inhibiting factor) [Source:HGNC Symbol;Acc:HGNC:7097] | effector | 1 | 1 | 0.516190532 |
| FOXA2 | forkhead box A2 [Source:HGNC Symbol;Acc:HGNC:5022] | transcription | 0 | 1 | 0.516320523 |
| TRIM74 | tripartite motif containing 74 [Source:HGNC Symbol;Acc:HGNC:17453] | regulator | 0 | 0 | 0.516366669 |
| DNAJA3 | DnaJ (Hsp40) homolog, subfamily A, member 3 [Source:HGNC Symbol;Acc:HGNC:11808] | regulator | 1 | 0 | 0.51700446 |
| TRIM49 | tripartite motif containing 49 [Source:HGNC Symbol;Acc:HGNC:13431] | regulator | 0 | 1 | 0.517234388 |
| DEFB107B | defensin, beta 107B [Source:HGNC Symbol;Acc:HGNC:31918] | effector | 0 | 0 | 0.517598658 |
| DEFB134 | defensin, beta 134 [Source:HGNC Symbol;Acc:HGNC:32399] | effector | 1 | 0 | 0.51776687 |
| MT2A | metallothionein 2A [Source:HGNC Symbol;Acc:HGNC:7406] | UC | 1 | 0 | 0.517780964 |
| TNIP3 | TNFAIP3 interacting protein 3 [Source:HGNC Symbol;Acc:HGNC:19315] | regulator | 0 | 1 | 0.518236166 |
| FGF6 | fibroblast growth factor 6 [Source:HGNC Symbol;Acc:HGNC:3684] | effector | 1 | 0 | 0.518547553 |
| DHX33 | DEAH (Asp-Glu-Ala-His) box polypeptide 33 [Source:HGNC Symbol;Acc:HGNC:16718] | receptor | 0 | 0 | 0.518579729 |
| PTK6 | protein tyrosine kinase 6 [Source:HGNC Symbol;Acc:HGNC:9617] | regulator | 1 | 0 | 0.518656561 |

| | | | | | |
|---|---|---|---|---|---|
| MAFB | v-maf avian musculoaponeurotic fibrosarcoma oncogene homolog B [Source:HGNC Symbol;Acc:HGNC:6408] | transcription | 0 | 1 | 0.518978888 |
| IL36G | interleukin 36, gamma [Source:HGNC Symbol;Acc:HGNC:15741] | effector | 1 | 0 | 0.51929839 |
| VLDLR | very low density lipoprotein receptor [Source:HGNC Symbol;Acc:HGNC:12698] | receptor | 0 | 1 | 0.519412614 |
| IL1A | interleukin 1, alpha [Source:HGNC Symbol;Acc:HGNC:5991] | effector | 0 | 0 | 0.519627452 |
| IL2RB | interleukin 2 receptor, beta [Source:HGNC Symbol;Acc:HGNC:6009] | secondary receptor | 0 | 1 | 0.519736206 |
| TRIM55 | tripartite motif containing 55 [Source:HGNC Symbol;Acc:HGNC:14215] | regulator | 0 | 1 | 0.519748853 |
| IFNGR2 | interferon gamma receptor 2 (interferon gamma transducer 1) [Source:HGNC Symbol;Acc:HGNC:5440] | secondary receptor | 1 | 1 | 0.51991665 |
| TYK2 | tyrosine kinase 2 [Source:HGNC Symbol;Acc:HGNC:12440] | signal transducer | 1 | 1 | 0.520423069 |
| KL | klotho [Source:HGNC Symbol;Acc:HGNC:6344] | secondary receptor | 1 | 0 | 0.520511994 |
| FGA | fibrinogen alpha chain [Source:HGNC Symbol;Acc:HGNC:3661] | effector | 1 | 0 | 0.521063863 |
| NLRP6 | NLR family, pyrin domain containing 6 [Source:HGNC Symbol;Acc:HGNC:22944] | receptor | 0 | 1 | 0.521218083 |
| POLR3C | polymerase (RNA) III (DNA directed) polypeptide C (62kD) [Source:HGNC Symbol;Acc:HGNC:30076] | receptor | 1 | 0 | 0.521255178 |
| NOD1 | nucleotide-binding oligomerization domain containing 1 [Source:HGNC Symbol;Acc:HGNC:16390] | receptor | 1 | 1 | 0.521989867 |
| RNF135 | ring finger protein 135 [Source:HGNC Symbol;Acc:HGNC:21158] | regulator | 1 | 1 | 0.522219004 |
| IFNL1 | interferon, lambda 1 [Source:HGNC Symbol;Acc:HGNC:18363] | effector | 0 | 0 | 0.522740528 |
| IL27 | interleukin 27 [Source:HGNC Symbol;Acc:HGNC:19157] | effector | 1 | 1 | 0.522745439 |
| C1S | complement component 1, s subcomponent [Source:HGNC Symbol;Acc:HGNC:1247] | signal transducer | 1 | 1 | 0.522853614 |
| IL25 | interleukin 25 [Source:HGNC Symbol;Acc:HGNC:13765] | effector | 0 | 1 | 0.522975118 |
| SLC2A11 | solute carrier family 2 (facilitated glucose transporter), member 11 [Source:HGNC Symbol;Acc:HGNC:14239] | UC | 0 | 0 | 0.523108135 |
| KIAA0226 | KIAA0226 [Source:HGNC Symbol;Acc:HGNC:28991] | regulator | 0 | 1 | 0.523567665 |
| IFNA17 | interferon, alpha 17 [Source:HGNC Symbol;Acc:HGNC:5422] | effector | 1 | 0 | 0.523627099 |
| CTSB | cathepsin B [Source:HGNC Symbol;Acc:HGNC:2527] | accessory molecule | 1 | 1 | 0.523878476 |
| SDC4 | syndecan 4 [Source:HGNC Symbol;Acc:HGNC:10661] | effector | 0 | 1 | 0.52406079 |
| STAT2 | signal transducer and activator of transcription 2, 113kDa [Source:HGNC Symbol;Acc:HGNC:11363] | transcription | 1 | 1 | 0.524855819 |
| GBP1 | guanylate binding protein 1, interferon-inducible [Source:HGNC Symbol;Acc:HGNC:4182] | effector | 1 | 0 | 0.525681437 |

| | | | | | |
|---|---|---|---|---|---|
| GLI1 | GLI family zinc finger 1 [Source:HGNC Symbol;Acc:HGNC:4317] | regulator | 0 | 1 | 0.52620915 |
| LAT | linker for activation of T cells [Source:HGNC Symbol;Acc:HGNC:18874] | signal transducer | 1 | 1 | 0.527174423 |
| LTF | lactotransferrin [Source:HGNC Symbol;Acc:HGNC:6720] | effector | 1 | 0 | 0.528087394 |
| POLR1C | polymerase (RNA) I polypeptide C, 30kDa [Source:HGNC Symbol;Acc:HGNC:20194] | receptor | 1 | 0 | 0.528129291 |
| PANX1 | pannexin 1 [Source:HGNC Symbol;Acc:HGNC:8599] | UC | 1 | 0 | 0.528403413 |
| FGF23 | fibroblast growth factor 23 [Source:HGNC Symbol;Acc:HGNC:3680] | effector | 1 | 0 | 0.528638492 |
| COLEC12 | collectin sub-family member 12 [Source:HGNC Symbol;Acc:HGNC:16016] | receptor | 1 | 0 | 0.528763986 |
| NFKBIB | nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, beta [Source:HGNC Symbol;Acc:HGNC:7798] | signal transducer | 1 | 1 | 0.528884381 |
| ERBB3 | erb-b2 receptor tyrosine kinase 3 [Source:HGNC Symbol;Acc:HGNC:3431] | receptor | 1 | 0 | 0.529075275 |
| SCAF11 | SR-related CTD-associated factor 11 [Source:HGNC Symbol;Acc:HGNC:10784] | signal transducer | 0 | 1 | 0.529346562 |
| POLR3E | polymerase (RNA) III (DNA directed) polypeptide E (80kD) [Source:HGNC Symbol;Acc:HGNC:30347] | receptor | 1 | 0 | 0.529366346 |
| JAM3 | junctional adhesion molecule 3 [Source:HGNC Symbol;Acc:HGNC:15532] | regulator | 0 | 1 | 0.530081221 |
| ACAP1 | ArfGAP with coiled-coil, ankyrin repeat and PH domains 1 [Source:HGNC Symbol;Acc:HGNC:16467] | UC | 0 | 1 | 0.530292386 |
| CAMP | cathelicidin antimicrobial peptide [Source:HGNC Symbol;Acc:HGNC:1472] | effector | 1 | 1 | 0.530312179 |
| SREBF2 | sterol regulatory element binding transcription factor 2 [Source:HGNC Symbol;Acc:HGNC:11290] | transcription | 0 | 1 | 0.530342904 |
| TICAM2 | toll-like receptor adaptor molecule 2 [Source:HGNC Symbol;Acc:HGNC:21354] | adaptor | 1 | 1 | 0.530694635 |
| IFI30 | interferon, gamma-inducible protein 30 [Source:HGNC Symbol;Acc:HGNC:5398] | accessory molecule | 1 | 0 | 0.530757078 |
| RGS2 | regulator of G-protein signaling 2 [Source:HGNC Symbol;Acc:HGNC:9998] | regulator | 0 | 1 | 0.531047306 |
| DEFB118 | defensin, beta 118 [Source:HGNC Symbol;Acc:HGNC:16196] | effector | 1 | 0 | 0.531192862 |
| SIVA1 | SIVA1, apoptosis-inducing factor [Source:HGNC Symbol;Acc:HGNC:17712] | regulator | 0 | 1 | 0.531291416 |
| DEFB135 | defensin, beta 135 [Source:HGNC Symbol;Acc:HGNC:32400] | effector | 1 | 0 | 0.531648254 |
| LRRFIP2 | leucine rich repeat (in FLII) interacting protein 2 [Source:HGNC Symbol;Acc:HGNC:6703] | regulator | 0 | 1 | 0.531914448 |
| TRIM43 | tripartite motif containing 43 [Source:HGNC Symbol;Acc:HGNC:19015] | regulator | 0 | 0 | 0.53193314 |
| BCAR1 | breast cancer anti-estrogen resistance 1 [Source:HGNC Symbol;Acc:HGNC:971] | UC | 0 | 1 | 0.532223958 |
| LYZ | lysozyme [Source:HGNC Symbol;Acc:HGNC:6740] | effector | 0 | 0 | 0.532887613 |
| LY75 | lymphocyte antigen 75 [Source:HGNC Symbol;Acc:HGNC:6729] | receptor | 0 | 0 | 0.532985888 |
| MMP9 | matrix metallopeptidase 9 (gelatinase B, 92kDa gelatinase, 92kDa type IV collagenase) [Source:HGNC Symbol;Acc:HGNC:7176] | accessory molecule | 0 | 1 | 0.533081842 |

| | | | | | |
|---|---|---|---|---|---|
| SERPINB9 | serpin peptidase inhibitor, clade B (ovalbumin), member 9 [Source:HGNC Symbol;Acc:HGNC:8955] | effector | 0 | 1 | 0.533133413 |
| LY96 | lymphocyte antigen 96 [Source:HGNC Symbol;Acc:HGNC:17156] | receptor | 1 | 1 | 0.533414876 |
| STX11 | syntaxin 11 [Source:HGNC Symbol;Acc:HGNC:11429] | accessory molecule | 1 | 0 | 0.533591781 |
| CCL7 | chemokine (C-C motif) ligand 7 [Source:HGNC Symbol;Acc:HGNC:10634] | effector | 0 | 0 | 0.533622203 |
| KLRF1 | killer cell lectin-like receptor subfamily F, member 1 [Source:HGNC Symbol;Acc:HGNC:13342] | receptor | 0 | 0 | 0.533696913 |
| CLEC10A | C-type lectin domain family 10, member A [Source:HGNC Symbol;Acc:HGNC:16916] | receptor | 1 | 0 | 0.533849954 |
| CLEC1A | C-type lectin domain family 1, member A [Source:HGNC Symbol;Acc:HGNC:24355] | receptor | 0 | 0 | 0.53424586 |
| DUOX1 | dual oxidase 1 [Source:HGNC Symbol;Acc:HGNC:3062] | UC | 0 | 1 | 0.534531293 |
| SLAMF1 | signaling lymphocytic activation molecule family member 1 [Source:HGNC Symbol;Acc:HGNC:10903] | receptor | 0 | 1 | 0.535403447 |
| ERAP1 | endoplasmic reticulum aminopeptidase 1 [Source:HGNC Symbol;Acc:HGNC:18173] | UC | 0 | 1 | 0.535449928 |
| CLEC12B | C-type lectin domain family 12, member B [Source:HGNC Symbol;Acc:HGNC:31966] | receptor | 0 | 0 | 0.535995934 |
| TRIM15 | tripartite motif containing 15 [Source:HGNC Symbol;Acc:HGNC:16284] | regulator | 1 | 1 | 0.537129917 |
| RGMB | repulsive guidance molecule family member b [Source:HGNC Symbol;Acc:HGNC:26896] | signal transducer | 0 | 1 | 0.537181496 |
| CD209 | CD209 molecule [Source:HGNC Symbol;Acc:HGNC:1641] | receptor | 1 | 1 | 0.537649382 |
| C1QTNF3 | C1q and tumor necrosis factor related protein 3 [Source:HGNC Symbol;Acc:HGNC:14326] | regulator | 0 | 0 | 0.537965574 |
| CASP10 | caspase 10, apoptosis-related cysteine peptidase [Source:HGNC Symbol;Acc:HGNC:1500] | signal transducer | 1 | 1 | 0.538082532 |
| IFIT2 | interferon-induced protein with tetratricopeptide repeats 2 [Source:HGNC Symbol;Acc:HGNC:5409] | receptor | 1 | 1 | 0.538383025 |
| CALCA | calcitonin-related polypeptide alpha [Source:HGNC Symbol;Acc:HGNC:1437] | effector | 0 | 1 | 0.538426123 |
| CXCL5 | chemokine (C-X-C motif) ligand 5 [Source:HGNC Symbol;Acc:HGNC:10642] | effector | 0 | 0 | 0.539049996 |
| NRIP1 | nuclear receptor interacting protein 1 [Source:HGNC Symbol;Acc:HGNC:8001] | transcription | 0 | 1 | 0.539654778 |
| C8G | complement component 8, gamma polypeptide [Source:HGNC Symbol;Acc:HGNC:1354] | signal transducer | 1 | 1 | 0.539775989 |
| PTK2B | protein tyrosine kinase 2 beta [Source:HGNC Symbol;Acc:HGNC:9612] | signal transducer | 1 | 1 | 0.540367482 |
| PIK3R3 | phosphoinositide-3-kinase, regulatory subunit 3 (gamma) [Source:HGNC Symbol;Acc:HGNC:8981] | signal transducer | 0 | 0 | 0.540942099 |
| F12 | coagulation factor XII (Hageman factor) [Source:HGNC Symbol;Acc:HGNC:3530] | signal transducer | 1 | 0 | 0.541183914 |
| DEFB133 | defensin, beta 133 [Source:HGNC Symbol;Acc:HGNC:31331] | effector | 0 | 0 | 0.541410871 |

| | | | | | |
|---|---|---|---|---|---|
| CCL4 | chemokine (C-C motif) ligand 4 [Source:HGNC Symbol;Acc:HGNC:10630] | effector | 0 | 0 | 0.542023068 |
| CD14 | CD14 molecule [Source:HGNC Symbol;Acc:HGNC:1628] | receptor | 1 | 1 | 0.542971764 |
| SFTPA1 | surfactant protein A1 [Source:HGNC Symbol;Acc:HGNC:10798] | receptor | 0 | 0 | 0.543082865 |
| IRAK3 | interleukin-1 receptor-associated kinase 3 [Source:HGNC Symbol;Acc:HGNC:17020] | signal transducer | 0 | 1 | 0.543198199 |
| LY86 | lymphocyte antigen 86 [Source:HGNC Symbol;Acc:HGNC:16837] | adaptor | 1 | 1 | 0.543736151 |
| CASP5 | caspase 5, apoptosis-related cysteine peptidase [Source:HGNC Symbol;Acc:HGNC:1506] | signal transducer | 0 | 0 | 0.54421279 |
| SIGLEC10 | sialic acid binding Ig-like lectin 10 [Source:HGNC Symbol;Acc:HGNC:15620] | receptor | 0 | 1 | 0.544525794 |
| CD86 | CD86 molecule [Source:HGNC Symbol;Acc:HGNC:1705] | effector | 1 | 0 | 0.544797066 |
| LSP1 | lymphocyte-specific protein 1 [Source:HGNC Symbol;Acc:HGNC:6707] | adaptor | 0 | 0 | 0.545219594 |
| VCAM1 | vascular cell adhesion molecule 1 [Source:HGNC Symbol;Acc:HGNC:12663] | UC | 1 | 0 | 0.545278841 |
| TRIM66 | tripartite motif containing 66 [Source:HGNC Symbol;Acc:HGNC:29005] | regulator | 0 | 1 | 0.545288869 |
| NFE2L2 | nuclear factor, erythroid 2-like 2 [Source:HGNC Symbol;Acc:HGNC:7782] | signal transducer | 0 | 1 | 0.545456166 |
| PADI4 | peptidyl arginine deiminase, type IV [Source:HGNC Symbol;Acc:HGNC:18368] | UC | 1 | 0 | 0.545896742 |
| SOCS2 | suppressor of cytokine signaling 2 [Source:HGNC Symbol;Acc:HGNC:19382] | effector | 0 | 1 | 0.545914916 |
| CDKN2A | cyclin-dependent kinase inhibitor 2A [Source:HGNC Symbol;Acc:HGNC:1787] | regulator | 0 | 1 | 0.546341486 |
| C5 | complement component 5 [Source:HGNC Symbol;Acc:HGNC:1331] | signal transducer | 1 | 1 | 0.546626927 |
| TRIM59 | tripartite motif containing 59 [Source:HGNC Symbol;Acc:HGNC:30834] | regulator | 0 | 0 | 0.546929057 |
| TICAM1 | toll-like receptor adaptor molecule 1 [Source:HGNC Symbol;Acc:HGNC:18348] | adaptor | 1 | 1 | 0.546962974 |
| DEFB116 | defensin, beta 116 [Source:HGNC Symbol;Acc:HGNC:18097] | effector | 1 | 0 | 0.547368698 |
| TRIM10 | tripartite motif containing 10 [Source:HGNC Symbol;Acc:HGNC:10072] | regulator | 1 | 0 | 0.547953521 |
| KLRG2 | killer cell lectin-like receptor subfamily G, member 2 [Source:HGNC Symbol;Acc:HGNC:24778] | receptor | 0 | 0 | 0.548127268 |
| MFGE8 | milk fat globule-EGF factor 8 protein [Source:HGNC Symbol;Acc:HGNC:7036] | regulator | 0 | 0 | 0.548141091 |
| TRIM31 | tripartite motif containing 31 [Source:HGNC Symbol;Acc:HGNC:16289] | regulator | 1 | 0 | 0.548769767 |
| MS4A8 | membrane-spanning 4-domains, subfamily A, member 8 [Source:HGNC Symbol;Acc:HGNC:13380] | effector | 0 | 0 | 0.548977415 |
| NFKBIE | nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, epsilon [Source:HGNC Symbol;Acc:HGNC:7799] | signal transducer | 0 | 1 | 0.549113777 |
| MB21D1 | Mab-21 domain containing 1 [Source:HGNC Symbol;Acc:HGNC:21367] | receptor | 1 | 1 | 0.549632713 |
| HBEGF | heparin-binding EGF-like growth factor [Source:HGNC Symbol;Acc:HGNC:3059] | effector | 1 | 0 | 0.549686031 |

| | | | | | |
|---|---|---|---|---|---|
| SIGLEC11 | sialic acid binding Ig-like lectin 11 [Source:HGNC Symbol;Acc:HGNC:15622] | receptor | 0 | 1 | 0.549875529 |
| TRAF5 | TNF receptor-associated factor 5 [Source:HGNC Symbol;Acc:HGNC:12035] | signal transducer | 0 | 1 | 0.550073377 |
| CLEC9A | C-type lectin domain family 9, member A [Source:HGNC Symbol;Acc:HGNC:26705] | receptor | 0 | 1 | 0.550440351 |
| IFNA6 | interferon, alpha 6 [Source:HGNC Symbol;Acc:HGNC:5427] | effector | 1 | 0 | 0.551228891 |
| NUP153 | nucleoporin 153kDa [Source:HGNC Symbol;Acc:HGNC:8062] | accessory molecule | 0 | 1 | 0.55128802 |
| IL13 | interleukin 13 [Source:HGNC Symbol;Acc:HGNC:5973] | effector | 0 | 1 | 0.551382735 |
| MSRB1 | methionine sulfoxide reductase B1 [Source:HGNC Symbol;Acc:HGNC:14133] | regulator | 1 | 0 | 0.551968713 |
| CD1D | CD1d molecule [Source:HGNC Symbol;Acc:HGNC:1637] | UC | 1 | 1 | 0.552081085 |
| ADAMTS13 | ADAM metallopeptidase with thrombospondin type 1 motif, 13 [Source:HGNC Symbol;Acc:HGNC:1366] | UC | 1 | 0 | 0.5520831 |
| CASP1 | caspase 1, apoptosis-related cysteine peptidase [Source:HGNC Symbol;Acc:HGNC:1499] | signal transducer | 1 | 1 | 0.552734547 |
| CCL13 | chemokine (C-C motif) ligand 13 [Source:HGNC Symbol;Acc:HGNC:10611] | effector | 0 | 0 | 0.552871547 |
| XAF1 | XIAP associated factor 1 [Source:HGNC Symbol;Acc:HGNC:30932] | regulator | 1 | 0 | 0.553034907 |
| LYST | lysosomal trafficking regulator [Source:HGNC Symbol;Acc:HGNC:1968] | accessory molecule | 1 | 0 | 0.553064298 |
| TRIM68 | tripartite motif containing 68 [Source:HGNC Symbol;Acc:HGNC:21161] | regulator | 0 | 0 | 0.553187991 |
| IL36RN | interleukin 36 receptor antagonist [Source:HGNC Symbol;Acc:HGNC:15561] | regulator | 1 | 0 | 0.553408702 |
| CLEC5A | C-type lectin domain family 5, member A [Source:HGNC Symbol;Acc:HGNC:2054] | receptor | 1 | 0 | 0.55399466 |
| SIGLEC7 | sialic acid binding Ig-like lectin 7 [Source:HGNC Symbol;Acc:HGNC:10876] | receptor | 0 | 1 | 0.55415514 |
| COLEC10 | collectin sub-family member 10 (C-type lectin) [Source:HGNC Symbol;Acc:HGNC:2220] | receptor | 0 | 0 | 0.554905914 |
| PRF1 | perforin 1 (pore forming protein) [Source:HGNC Symbol;Acc:HGNC:9360] | effector | 0 | 0 | 0.554954908 |
| NR1H4 | nuclear receptor subfamily 1, group H, member 4 [Source:HGNC Symbol;Acc:HGNC:7967] | transcription | 0 | 1 | 0.555002019 |
| LILRA5 | leukocyte immunoglobulin-like receptor, subfamily A (with TM domain), member 5 [Source:HGNC Symbol;Acc:HGNC:16309] | receptor | 1 | 0 | 0.555536503 |
| CLEC17A | C-type lectin domain family 17, member A [Source:HGNC Symbol;Acc:HGNC:34520] | receptor | 0 | 0 | 0.556093454 |
| NLRP1 | NLR family, pyrin domain containing 1 [Source:HGNC Symbol;Acc:HGNC:14374] | receptor | 1 | 1 | 0.556725649 |
| GRN | granulin [Source:HGNC Symbol;Acc:HGNC:4601] | effector | 0 | 1 | 0.557157585 |
| CD22 | CD22 molecule [Source:HGNC Symbol;Acc:HGNC:1643] | regulator | 0 | 1 | 0.55743144 |
| SELE | selectin E [Source:HGNC Symbol;Acc:HGNC:10718] | effector | 0 | 1 | 0.558084552 |
| CCL11 | chemokine (C-C motif) ligand 11 [Source:HGNC Symbol;Acc:HGNC:10610] | effector | 0 | 0 | 0.559210746 |

| | | | | | |
|---|---|---|---|---|---|
| ATG12 | autophagy related 12 [Source:HGNC Symbol;Acc:HGNC:588] | regulator | 1 | 1 | 0.560028894 |
| RSAD2 | radical S-adenosyl methionine domain containing 2 [Source:HGNC Symbol;Acc:HGNC:30908] | effector | 1 | 1 | 0.560087496 |
| CD97 | CD97 molecule [Source:HGNC Symbol;Acc:HGNC:1711] | receptor | 0 | 1 | 0.560342048 |
| IFIT1 | interferon-induced protein with tetratricopeptide repeats 1 [Source:HGNC Symbol;Acc:HGNC:5407] | receptor | 1 | 1 | 0.560352199 |
| IFNLR1 | interferon, lambda receptor 1 [Source:HGNC Symbol;Acc:HGNC:18584] | secondary receptor | 1 | 0 | 0.560784357 |
| BTN3A1 | butyrophilin, subfamily 3, member A1 [Source:HGNC Symbol;Acc:HGNC:1138] | receptor | 0 | 1 | 0.560908761 |
| SLAMF8 | SLAM family member 8 [Source:HGNC Symbol;Acc:HGNC:21391] | receptor | 0 | 1 | 0.561073546 |
| XDH | xanthine dehydrogenase [Source:HGNC Symbol;Acc:HGNC:12805] | signal transducer | 0 | 1 | 0.561739114 |
| CD300E | CD300e molecule [Source:HGNC Symbol;Acc:HGNC:28874] | regulator | 1 | 1 | 0.56208099 |
| APAF1 | apoptotic peptidase activating factor 1 [Source:HGNC Symbol;Acc:HGNC:576] | signal transducer | 0 | 0 | 0.562269862 |
| IFNA5 | interferon, alpha 5 [Source:HGNC Symbol;Acc:HGNC:5426] | effector | 1 | 0 | 0.562777767 |
| PTX3 | pentraxin 3, long [Source:HGNC Symbol;Acc:HGNC:9692] | receptor | 1 | 1 | 0.563089187 |
| SRMS | src-related kinase lacking C-terminal regulatory tyrosine and N-terminal myristylation sites [Source:HGNC Symbol;Acc:HGNC:11298] | regulator | 1 | 0 | 0.563092864 |
| DEFA1 | defensin, alpha 1 [Source:HGNC Symbol;Acc:HGNC:2761] | effector | 1 | 1 | 0.563247106 |
| KLRF2 | killer cell lectin-like receptor subfamily F, member 2 [Source:HGNC Symbol;Acc:HGNC:37646] | receptor | 1 | 0 | 0.563251013 |
| ZFPM2 | zinc finger protein, FOG family member 2 [Source:HGNC Symbol;Acc:HGNC:16700] | regulator | 0 | 1 | 0.563383575 |
| IRG1 | immunoresponsive 1 homolog (mouse) [Source:HGNC Symbol;Acc:HGNC:33904] | regulator | 1 | 0 | 0.563642 |
| MMP7 | matrix metallopeptidase 7 (matrilysin, uterine) [Source:HGNC Symbol;Acc:HGNC:7174] | accessory molecule | 0 | 1 | 0.563779829 |
| SRPK1 | SRSF protein kinase 1 [Source:HGNC Symbol;Acc:HGNC:11305] | signal transducer | 1 | 0 | 0.565229889 |
| FGF5 | fibroblast growth factor 5 [Source:HGNC Symbol;Acc:HGNC:3683] | effector | 1 | 0 | 0.56529186 |
| CD44 | CD44 molecule (Indian blood group) [Source:HGNC Symbol;Acc:HGNC:1681] | receptor | 1 | 0 | 0.565361193 |
| KLB | klotho beta [Source:HGNC Symbol;Acc:HGNC:15527] | secondary receptor | 1 | 0 | 0.565884447 |
| TREX1 | three prime repair exonuclease 1 [Source:HGNC Symbol;Acc:HGNC:12269] | receptor | 1 | 0 | 0.566158975 |
| ATF4 | activating transcription factor 4 [Source:HGNC Symbol;Acc:HGNC:786] | transcription | 0 | 1 | 0.566798829 |
| TLR6 | toll-like receptor 6 [Source:HGNC Symbol;Acc:HGNC:16711] | receptor | 1 | 1 | 0.567265525 |
| SREBF1 | sterol regulatory element binding transcription factor 1 [Source:HGNC Symbol;Acc:HGNC:11289] | transcription | 0 | 1 | 0.567669688 |

| | | | | | |
|---|---|---|---|---|---|
| AIMP1 | aminoacyl tRNA synthetase complex-interacting multifunctional protein 1 [Source:HGNC Symbol;Acc:HGNC:10648] | signal transducer | 0 | 1 | 0.568072155 |
| PLG | plasminogen [Source:HGNC Symbol;Acc:HGNC:9071] | regulator | 0 | 1 | 0.568269005 |
| BTN3A2 | butyrophilin, subfamily 3, member A2 [Source:HGNC Symbol;Acc:HGNC:1139] | receptor | 0 | 1 | 0.568864854 |
| TRAFD1 | TRAF-type zinc finger domain containing 1 [Source:HGNC Symbol;Acc:HGNC:24808] | regulator | 0 | 1 | 0.568908824 |
| SLC30A8 | solute carrier family 30 (zinc transporter), member 8 [Source:HGNC Symbol;Acc:HGNC:20303] | UC | 1 | 0 | 0.568967875 |
| SIGLEC9 | sialic acid binding Ig-like lectin 9 [Source:HGNC Symbol;Acc:HGNC:10878] | receptor | 0 | 1 | 0.570284457 |
| CLEC18C | C-type lectin domain family 18, member C [Source:HGNC Symbol;Acc:HGNC:28538] | receptor | 0 | 0 | 0.57029428 |
| DEFB124 | defensin, beta 124 [Source:HGNC Symbol;Acc:HGNC:18104] | effector | 1 | 0 | 0.570336856 |
| USP17L2 | ubiquitin specific peptidase 17-like family member 2 [Source:HGNC Symbol;Acc:HGNC:34434] | regulator | 0 | 0 | 0.570862729 |
| CD80 | CD80 molecule [Source:HGNC Symbol;Acc:HGNC:1700] | effector | 1 | 0 | 0.571179622 |
| TRAF1 | TNF receptor-associated factor 1 [Source:HGNC Symbol;Acc:HGNC:12031] | signal transducer | 0 | 1 | 0.572226389 |
| KLRB1 | killer cell lectin-like receptor subfamily B, member 1 [Source:HGNC Symbol;Acc:HGNC:6373] | receptor | 0 | 0 | 0.572402558 |
| MEFV | Mediterranean fever [Source:HGNC Symbol;Acc:HGNC:6998] | regulator | 1 | 1 | 0.572695768 |
| MS4A2 | membrane-spanning 4-domains, subfamily A, member 2 [Source:HGNC Symbol;Acc:HGNC:7316] | receptor | 1 | 0 | 0.573003117 |
| CTSG | cathepsin G [Source:HGNC Symbol;Acc:HGNC:2532] | accessory molecule | 0 | 1 | 0.573038991 |
| C1QB | complement component 1, q subcomponent, B chain [Source:HGNC Symbol;Acc:HGNC:1242] | receptor | 1 | 1 | 0.573331774 |
| ACKR2 | atypical chemokine receptor 2 [Source:HGNC Symbol;Acc:HGNC:1565] | secondary receptor | 0 | 0 | 0.573586355 |
| ART1 | ADP-ribosyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:723] | UC | 1 | 0 | 0.573725261 |
| WNT9B | wingless-type MMTV integration site family, member 9B [Source:HGNC Symbol;Acc:HGNC:12779] | effector | 0 | 1 | 0.574126417 |
| GSTP1 | glutathione S-transferase pi 1 [Source:HGNC Symbol;Acc:HGNC:4638] | regulator | 0 | 1 | 0.574171674 |
| ULBP3 | UL16 binding protein 3 [Source:HGNC Symbol;Acc:HGNC:14895] | effector | 1 | 0 | 0.574672273 |
| CLEC4E | C-type lectin domain family 4, member E [Source:HGNC Symbol;Acc:HGNC:14555] | receptor | 0 | 1 | 0.576565572 |
| DEFB112 | defensin, beta 112 [Source:HGNC Symbol;Acc:HGNC:18093] | effector | 0 | 0 | 0.578327407 |
| CD207 | CD207 molecule, langerin [Source:HGNC Symbol;Acc:HGNC:17935] | receptor | 0 | 0 | 0.579140224 |
| GZMM | granzyme M (lymphocyte met-ase 1) [Source:HGNC Symbol;Acc:HGNC:4712] | signal transducer | 1 | 1 | 0.579895358 |
| MNDA | myeloid cell nuclear differentiation antigen [Source:HGNC Symbol;Acc:HGNC:7183] | receptor | 0 | 0 | 0.58074548 |
| EPOR | erythropoietin receptor [Source:HGNC Symbol;Acc:HGNC:3416] | secondary receptor | 0 | 1 | 0.581040933 |

| | | | | | |
|---|---|---|---|---|---|
| FFAR2 | free fatty acid receptor 2 [Source:HGNC Symbol;Acc:HGNC:4501] | UC | 0 | 1 | 0.581606951 |
| IL1R2 | interleukin 1 receptor, type II [Source:HGNC Symbol;Acc:HGNC:5994] | secondary receptor | 0 | 1 | 0.581983337 |
| AQP4 | aquaporin 4 [Source:HGNC Symbol;Acc:HGNC:637] | UC | 1 | 0 | 0.58362102 |
| UNC13D | unc-13 homolog D (C. elegans) [Source:HGNC Symbol;Acc:HGNC:23147] | accessory molecule | 1 | 0 | 0.584106614 |
| IFI27 | interferon, alpha-inducible protein 27 [Source:HGNC Symbol;Acc:HGNC:5397] | effector | 1 | 1 | 0.58455027 |
| DEFB127 | defensin, beta 127 [Source:HGNC Symbol;Acc:HGNC:16206] | effector | 1 | 0 | 0.585202953 |
| SAA1 | serum amyloid A1 [Source:HGNC Symbol;Acc:HGNC:10513] | effector | 1 | 0 | 0.585824447 |
| BDKRB2 | bradykinin receptor B2 [Source:HGNC Symbol;Acc:HGNC:1030] | receptor | 0 | 1 | 0.586123422 |
| TRIM16L | tripartite motif containing 16-like [Source:HGNC Symbol;Acc:HGNC:32670] | regulator | 0 | 0 | 0.586685972 |
| LGALS2 | lectin, galactoside-binding, soluble, 2 [Source:HGNC Symbol;Acc:HGNC:6562] | receptor | 0 | 1 | 0.587146544 |
| CYTIP | cytohesin 1 interacting protein [Source:HGNC Symbol;Acc:HGNC:9506] | UC | 0 | 1 | 0.588487236 |
| USP4 | ubiquitin specific peptidase 4 (proto-oncogene) [Source:HGNC Symbol;Acc:HGNC:12627] | regulator | 0 | 1 | 0.588937928 |
| IFI35 | interferon-induced protein 35 [Source:HGNC Symbol;Acc:HGNC:5399] | regulator | 1 | 0 | 0.589743532 |
| SIGIRR | single immunoglobulin and toll-interleukin 1 receptor (TIR) domain [Source:HGNC Symbol;Acc:HGNC:30575] | regulator | 0 | 1 | 0.590030592 |
| CHGA | chromogranin A (parathyroid secretory protein 1) [Source:HGNC Symbol;Acc:HGNC:1929] | UC | 0 | 1 | 0.590553856 |
| FGF22 | fibroblast growth factor 22 [Source:HGNC Symbol;Acc:HGNC:3679] | effector | 1 | 0 | 0.590706473 |
| F2RL3 | coagulation factor II (thrombin) receptor-like 3 [Source:HGNC Symbol;Acc:HGNC:3540] | secondary receptor | 0 | 1 | 0.590834359 |
| TNFRSF1B | tumor necrosis factor receptor superfamily, member 1B [Source:HGNC Symbol;Acc:HGNC:11917] | secondary receptor | 0 | 1 | 0.591146424 |
| CARD8 | caspase recruitment domain family, member 8 [Source:HGNC Symbol;Acc:HGNC:17057] | adaptor | 0 | 0 | 0.591307965 |
| LZTS1 | leucine zipper, putative tumor suppressor 1 [Source:HGNC Symbol;Acc:HGNC:13861] | UC | 1 | 0 | 0.591789255 |
| RNASEL | ribonuclease L (2',5'-oligoisoadenylate synthetase-dependent) [Source:HGNC Symbol;Acc:HGNC:10050] | signal transducer | 1 | 1 | 0.592486276 |
| C4BPA | complement component 4 binding protein, alpha [Source:HGNC Symbol;Acc:HGNC:1325] | regulator | 1 | 1 | 0.592696015 |
| IRS1 | insulin receptor substrate 1 [Source:HGNC Symbol;Acc:HGNC:6125] | effector | 1 | 0 | 0.592937493 |
| MASP2 | mannan-binding lectin serine peptidase 2 [Source:HGNC Symbol;Acc:HGNC:6902] | adaptor | 1 | 1 | 0.593263849 |
| IRGM | immunity-related GTPase family, M [Source:HGNC Symbol;Acc:HGNC:29597] | regulator | 1 | 0 | 0.594287394 |
| CFHR5 | complement factor H-related 5 [Source:HGNC Symbol;Acc:HGNC:24668] | regulator | 1 | 0 | 0.594293881 |
| TREML1 | triggering receptor expressed on myeloid cells-like 1 [Source:HGNC Symbol;Acc:HGNC:20434] | receptor | 1 | 0 | 0.594731858 |

| | | | | | |
|---|---|---|---|---|---|
| CLEC18A | C-type lectin domain family 18, member A [Source:HGNC Symbol;Acc:HGNC:30388] | receptor | 0 | 0 | 0.595679462 |
| TRIM42 | tripartite motif containing 42 [Source:HGNC Symbol;Acc:HGNC:19014] | regulator | 0 | 1 | 0.596125558 |
| LRRK2 | leucine-rich repeat kinase 2 [Source:HGNC Symbol;Acc:HGNC:18618] | signal transducer | 0 | 1 | 0.596354994 |
| CEACAM1 | carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein) [Source:HGNC Symbol;Acc:HGNC:1814] | regulator | 0 | 1 | 0.597231822 |
| USP2 | ubiquitin specific peptidase 2 [Source:HGNC Symbol;Acc:HGNC:12618] | regulator | 0 | 1 | 0.597461185 |
| PGLYRP3 | peptidoglycan recognition protein 3 [Source:HGNC Symbol;Acc:HGNC:30014] | receptor | 1 | 1 | 0.597796849 |
| IRAK4 | interleukin-1 receptor-associated kinase 4 [Source:HGNC Symbol;Acc:HGNC:17967] | signal transducer | 1 | 1 | 0.598013444 |
| IFNA14 | interferon, alpha 14 [Source:HGNC Symbol;Acc:HGNC:5420] | effector | 1 | 0 | 0.598436907 |
| IFIH1 | interferon induced with helicase C domain 1 [Source:HGNC Symbol;Acc:HGNC:18873] | receptor | 1 | 1 | 0.598659336 |
| HRH4 | histamine receptor H4 [Source:HGNC Symbol;Acc:HGNC:17383] | secondary receptor | 0 | 1 | 0.598781468 |
| TNRC6A | trinucleotide repeat containing 6A [Source:HGNC Symbol;Acc:HGNC:11969] | regulator | 1 | 0 | 0.598842131 |
| TRIM58 | tripartite motif containing 58 [Source:HGNC Symbol;Acc:HGNC:24150] | regulator | 0 | 1 | 0.59885491 |
| LY9 | lymphocyte antigen 9 [Source:HGNC Symbol;Acc:HGNC:6730] | UC | 0 | 1 | 0.599005795 |
| FGFR4 | fibroblast growth factor receptor 4 [Source:HGNC Symbol;Acc:HGNC:3691] | receptor | 1 | 0 | 0.599510119 |
| DEFA4 | defensin, alpha 4, corticostatin [Source:HGNC Symbol;Acc:HGNC:2763] | effector | 1 | 1 | 0.600383392 |
| UNC5CL | unc-5 homolog C (C. elegans)-like [Source:HGNC Symbol;Acc:HGNC:21203] | UC | 0 | 1 | 0.600926841 |
| NOX4 | NADPH oxidase 4 [Source:HGNC Symbol;Acc:HGNC:7891] | effector | 0 | 1 | 0.601815557 |
| PARK2 | parkin RBR E3 ubiquitin protein ligase [Source:HGNC Symbol;Acc:HGNC:8607] | accessory molecule | 0 | 1 | 0.602111746 |
| TRIM40 | tripartite motif containing 40 [Source:HGNC Symbol;Acc:HGNC:18736] | regulator | 0 | 0 | 0.602440104 |
| MAPK12 | mitogen-activated protein kinase 12 [Source:HGNC Symbol;Acc:HGNC:6874] | signal transducer | 0 | 0 | 0.602648619 |
| CD5L | CD5 molecule-like [Source:HGNC Symbol;Acc:HGNC:1690] | regulator | 0 | 1 | 0.602758329 |
| DEFB113 | defensin, beta 113 [Source:HGNC Symbol;Acc:HGNC:18094] | effector | 0 | 0 | 0.603313698 |
| DAK | dihydroxyacetone kinase 2 homolog (S. cerevisiae) [Source:HGNC Symbol;Acc:HGNC:24552] | regulator | 1 | 0 | 0.604034122 |
| NOX1 | NADPH oxidase 1 [Source:HGNC Symbol;Acc:HGNC:7889] | effector | 0 | 1 | 0.604108292 |
| DOK3 | docking protein 3 [Source:HGNC Symbol;Acc:HGNC:24583] | regulator | 0 | 1 | 0.60417878 |
| TNFRSF12A | tumor necrosis factor receptor superfamily, member 12A [Source:HGNC Symbol;Acc:HGNC:18152] | secondary receptor | 0 | 1 | 0.60442032 |

| | | | | | |
|---|---|---|---|---|---|
| LGALS3 | lectin, galactoside-binding, soluble, 3 [Source:HGNC Symbol;Acc:HGNC:6563] | receptor | 1 | 1 | 0.604462228 |
| YJEFN3 | YjeF N-terminal domain containing 3 [Source:HGNC Symbol;Acc:HGNC:24785] | UC | 0 | 1 | 0.605038952 |
| TRIM6 | tripartite motif containing 6 [Source:HGNC Symbol;Acc:HGNC:16277] | regulator | 0 | 1 | 0.605506727 |
| GBP7 | guanylate binding protein 7 [Source:HGNC Symbol;Acc:HGNC:29606] | effector | 0 | 0 | 0.605931666 |
| ITGAX | integrin, alpha X (complement component 3 receptor 4 subunit) [Source:HGNC Symbol;Acc:HGNC:6152] | receptor | 0 | 1 | 0.606054391 |
| RPS6KB2 | ribosomal protein S6 kinase, 70kDa, polypeptide 2 [Source:HGNC Symbol;Acc:HGNC:10437] | signal transducer | 1 | 0 | 0.607343451 |
| TLR10 | toll-like receptor 10 [Source:HGNC Symbol;Acc:HGNC:15634] | receptor | 1 | 1 | 0.607347179 |
| IL12B | interleukin 12B [Source:HGNC Symbol;Acc:HGNC:5970] | effector | 1 | 1 | 0.60747833 |
| IL34 | interleukin 34 [Source:HGNC Symbol;Acc:HGNC:28529] | effector | 1 | 0 | 0.608613636 |
| RAB27A | RAB27A, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:9766] | accessory molecule | 1 | 0 | 0.608711529 |
| C9 | complement component 9 [Source:HGNC Symbol;Acc:HGNC:1358] | signal transducer | 1 | 1 | 0.608980775 |
| ERBB2IP | erbb2 interacting protein [Source:HGNC Symbol;Acc:HGNC:15842] | regulator | 0 | 1 | 0.60961488 |
| PRSS3 | protease, serine, 3 [Source:HGNC Symbol;Acc:HGNC:9486] | effector | 1 | 0 | 0.610460549 |
| CASP6 | caspase 6, apoptosis-related cysteine peptidase [Source:HGNC Symbol;Acc:HGNC:1507] | signal transducer | 0 | 1 | 0.61091554 |
| BPIFA1 | BPI fold containing family A, member 1 [Source:HGNC Symbol;Acc:HGNC:15749] | effector | 1 | 0 | 0.61165564 |
| OASL | 2'-5'-oligoadenylate synthetase-like [Source:HGNC Symbol;Acc:HGNC:8090] | receptor | 1 | 0 | 0.612202132 |
| F2RL2 | coagulation factor II (thrombin) receptor-like 2 [Source:HGNC Symbol;Acc:HGNC:3539] | secondary receptor | 0 | 1 | 0.612839686 |
| DEFB110 | defensin, beta 110 locus [Source:HGNC Symbol;Acc:HGNC:18091] | effector | 0 | 0 | 0.613312606 |
| DEFB105B | defensin, beta 105B [Source:HGNC Symbol;Acc:HGNC:29930] | effector | 0 | 0 | 0.613432815 |
| ISG15 | ISG15 ubiquitin-like modifier [Source:HGNC Symbol;Acc:HGNC:4053] | regulator | 1 | 1 | 0.613465127 |
| HMOX1 | heme oxygenase (decycling) 1 [Source:HGNC Symbol;Acc:HGNC:5013] | signal transducer | 0 | 1 | 0.61351415 |
| MSR1 | macrophage scavenger receptor 1 [Source:HGNC Symbol;Acc:HGNC:7376] | receptor | 0 | 1 | 0.615217355 |
| HIST1H2BJ | histone cluster 1, H2bj [Source:HGNC Symbol;Acc:HGNC:4761] | UC | 1 | 0 | 0.615318349 |
| SIGLEC14 | sialic acid binding Ig-like lectin 14 [Source:HGNC Symbol;Acc:HGNC:32926] | receptor | 1 | 0 | 0.615363802 |
| PGLYRP2 | peptidoglycan recognition protein 2 [Source:HGNC Symbol;Acc:HGNC:30013] | receptor | 1 | 1 | 0.615440697 |
| CLEC1B | C-type lectin domain family 1, member B [Source:HGNC Symbol;Acc:HGNC:24356] | receptor | 0 | 1 | 0.615980896 |

| | | | | | |
|---|---|---|---|---|---|
| CSF1 | colony stimulating factor 1 (macrophage) [Source:HGNC Symbol;Acc:HGNC:2432] | effector | 1 | 0 | 0.616610007 |
| TRIM22 | tripartite motif containing 22 [Source:HGNC Symbol;Acc:HGNC:16379] | regulator | 0 | 1 | 0.616794133 |
| IFNA4 | interferon, alpha 4 [Source:HGNC Symbol;Acc:HGNC:5425] | effector | 1 | 0 | 0.616887227 |
| PHLPP2 | PH domain and leucine rich repeat protein phosphatase 2 [Source:HGNC Symbol;Acc:HGNC:29149] | signal transducer | 1 | 0 | 0.617611513 |
| LAIR1 | leukocyte-associated immunoglobulin-like receptor 1 [Source:HGNC Symbol;Acc:HGNC:6477] | receptor | 0 | 1 | 0.618045561 |
| CR2 | complement component (3d/Epstein Barr virus) receptor 2 [Source:HGNC Symbol;Acc:HGNC:2336] | secondary receptor | 1 | 1 | 0.618634172 |
| CCL3 | chemokine (C-C motif) ligand 3 [Source:HGNC Symbol;Acc:HGNC:10627] | effector | 1 | 0 | 0.618717617 |
| DEFB126 | defensin, beta 126 [Source:HGNC Symbol;Acc:HGNC:15900] | effector | 1 | 0 | 0.619117892 |
| CFI | complement factor I [Source:HGNC Symbol;Acc:HGNC:5394] | regulator | 1 | 0 | 0.619194286 |
| IL33 | interleukin 33 [Source:HGNC Symbol;Acc:HGNC:16028] | effector | 0 | 1 | 0.619844446 |
| ECSIT | ECSIT signalling integrator [Source:HGNC Symbol;Acc:HGNC:29548] | regulator | 1 | 1 | 0.619988032 |
| PPARGC1B | peroxisome proliferator-activated receptor gamma, coactivator 1 beta [Source:HGNC Symbol;Acc:HGNC:30022] | signal transducer | 0 | 1 | 0.620118225 |
| TIRAP | toll-interleukin 1 receptor (TIR) domain containing adaptor protein [Source:HGNC Symbol;Acc:HGNC:17192] | adaptor | 1 | 1 | 0.620700682 |
| CDKN1A | cyclin-dependent kinase inhibitor 1A (p21, Cip1) [Source:HGNC Symbol;Acc:HGNC:1784] | signal transducer | 1 | 0 | 0.620934914 |
| TRIM29 | tripartite motif containing 29 [Source:HGNC Symbol;Acc:HGNC:17274] | regulator | 0 | 0 | 0.621480638 |
| IFI16 | interferon, gamma-inducible protein 16 [Source:HGNC Symbol;Acc:HGNC:5395] | receptor | 1 | 1 | 0.621722528 |
| BPIFB3 | BPI fold containing family B, member 3 [Source:HGNC Symbol;Acc:HGNC:16178] | effector | 1 | 0 | 0.622135822 |
| OAS3 | 2'-5'-oligoadenylate synthetase 3, 100kDa [Source:HGNC Symbol;Acc:HGNC:8088] | receptor | 1 | 1 | 0.623775545 |
| DEFB129 | defensin, beta 129 [Source:HGNC Symbol;Acc:HGNC:16218] | effector | 1 | 0 | 0.623891045 |
| ULBP2 | UL16 binding protein 2 [Source:HGNC Symbol;Acc:HGNC:14894] | effector | 1 | 0 | 0.624362722 |
| DEFA1B | defensin, alpha 1B [Source:HGNC Symbol;Acc:HGNC:33596] | effector | 0 | 0 | 0.624654224 |
| SPP1 | secreted phosphoprotein 1 [Source:HGNC Symbol;Acc:HGNC:11255] | effector | 0 | 1 | 0.62526024 |
| ELP2 | elongator acetyltransferase complex subunit 2 [Source:HGNC Symbol;Acc:HGNC:18248] | regulator | 0 | 1 | 0.625833381 |
| MUC1 | mucin 1, cell surface associated [Source:HGNC Symbol;Acc:HGNC:7508] | regulator | 0 | 1 | 0.626589816 |
| LCN2 | lipocalin 2 [Source:HGNC Symbol;Acc:HGNC:6526] | UC | 1 | 0 | 0.627235085 |
| APOBEC3G | apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G [Source:HGNC Symbol;Acc:HGNC:17357] | effector | 1 | 1 | 0.627351168 |
| CXCL16 | chemokine (C-X-C motif) ligand 16 [Source:HGNC Symbol;Acc:HGNC:16642] | effector | 1 | 0 | 0.627863843 |

| | | | | | |
|---|---|---|---|---|---|
| SHARPIN | SHANK-associated RH domain interactor [Source:HGNC Symbol;Acc:HGNC:25321] | signal transducer | 0 | 1 | 0.627941977 |
| FCGR3A | Fc fragment of IgG, low affinity IIIa, receptor (CD16a) [Source:HGNC Symbol;Acc:HGNC:3619] | receptor | 1 | 0 | 0.627985215 |
| TYROBP | TYRO protein tyrosine kinase binding protein [Source:HGNC Symbol;Acc:HGNC:12449] | adaptor | 1 | 0 | 0.628049519 |
| CNKSR1 | connector enhancer of kinase suppressor of Ras 1 [Source:HGNC Symbol;Acc:HGNC:19700] | adaptor | 0 | 0 | 0.628624282 |
| DMBT1 | deleted in malignant brain tumors 1 [Source:HGNC Symbol;Acc:HGNC:2926] | effector | 1 | 1 | 0.629266861 |
| PKLR | pyruvate kinase, liver and RBC [Source:HGNC Symbol;Acc:HGNC:9020] | receptor | 0 | 1 | 0.63064263 |
| CLEC4M | C-type lectin domain family 4, member M [Source:HGNC Symbol;Acc:HGNC:13523] | receptor | 1 | 0 | 0.631290969 |
| NLRX1 | NLR family member X1 [Source:HGNC Symbol;Acc:HGNC:29890] | signal transducer | 1 | 1 | 0.631568655 |
| PGLYRP4 | peptidoglycan recognition protein 4 [Source:HGNC Symbol;Acc:HGNC:30015] | receptor | 1 | 1 | 0.632798073 |
| APOBEC3F | apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3F [Source:HGNC Symbol;Acc:HGNC:17356] | effector | 1 | 0 | 0.633404759 |
| CD300A | CD300a molecule [Source:HGNC Symbol;Acc:HGNC:19319] | regulator | 0 | 1 | 0.633832561 |
| DEFB114 | defensin, beta 114 [Source:HGNC Symbol;Acc:HGNC:18095] | effector | 0 | 0 | 0.63488599 |
| SIGLEC6 | sialic acid binding Ig-like lectin 6 [Source:HGNC Symbol;Acc:HGNC:10875] | receptor | 0 | 1 | 0.635066638 |
| DEFA5 | defensin, alpha 5, Paneth cell-specific [Source:HGNC Symbol;Acc:HGNC:2764] | effector | 1 | 1 | 0.635109651 |
| CD58 | CD58 molecule [Source:HGNC Symbol;Acc:HGNC:1688] | effector | 1 | 0 | 0.635456905 |
| CRYAB | crystallin, alpha B [Source:HGNC Symbol;Acc:HGNC:2389] | UC | 0 | 0 | 0.6359701 |
| ZC3HAV1 | zinc finger CCCH-type, antiviral 1 [Source:HGNC Symbol;Acc:HGNC:23721] | receptor | 1 | 1 | 0.636104757 |
| C5AR2 | complement component 5a receptor 2 [Source:HGNC Symbol;Acc:HGNC:4527] | secondary receptor | 0 | 0 | 0.636298092 |
| KLRD1 | killer cell lectin-like receptor subfamily D, member 1 [Source:HGNC Symbol;Acc:HGNC:6378] | receptor | 1 | 0 | 0.636482633 |
| CLEC2D | C-type lectin domain family 2, member D [Source:HGNC Symbol;Acc:HGNC:14351] | receptor | 0 | 0 | 0.63659717 |
| TRIM45 | tripartite motif containing 45 [Source:HGNC Symbol;Acc:HGNC:19018] | regulator | 0 | 1 | 0.636735435 |
| APOA4 | apolipoprotein A-IV [Source:HGNC Symbol;Acc:HGNC:602] | regulator | 1 | 0 | 0.637409168 |
| TNIP1 | TNFAIP3 interacting protein 1 [Source:HGNC Symbol;Acc:HGNC:16903] | regulator | 0 | 1 | 0.637478321 |
| CHAT | choline O-acetyltransferase [Source:HGNC Symbol;Acc:HGNC:1912] | UC | 0 | 0 | 0.637494141 |
| APOBEC3C | apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3C [Source:HGNC Symbol;Acc:HGNC:17353] | effector | 1 | 0 | 0.637819313 |
| CLEC6A | C-type lectin domain family 6, member A [Source:HGNC Symbol;Acc:HGNC:14556] | receptor | 1 | 1 | 0.638429368 |
| FCN3 | ficolin (collagen/fibrinogen domain containing) 3 [Source:HGNC Symbol;Acc:HGNC:3625] | receptor | 1 | 1 | 0.641938076 |

| | | | | | |
|---|---|---|---|---|---|
| TRIM60 | tripartite motif containing 60 [Source:HGNC Symbol;Acc:HGNC:21162] | regulator | 0 | 1 | 0.641956421 |
| IL17F | interleukin 17F [Source:HGNC Symbol;Acc:HGNC:16404] | effector | 0 | 1 | 0.642684373 |
| RETNLB | resistin like beta [Source:HGNC Symbol;Acc:HGNC:20388] | effector | 0 | 1 | 0.642747767 |
| TRIM34 | tripartite motif containing 34 [Source:HGNC Symbol;Acc:HGNC:10063] | regulator | 0 | 0 | 0.643085602 |
| IL17RE | interleukin 17 receptor E [Source:HGNC Symbol;Acc:HGNC:18439] | secondary receptor | 0 | 1 | 0.643434658 |
| ADIPOQ | adiponectin, C1Q and collagen domain containing [Source:HGNC Symbol;Acc:HGNC:13633] | regulator | 0 | 1 | 0.644336563 |
| CLEC12A | C-type lectin domain family 12, member A [Source:HGNC Symbol;Acc:HGNC:31713] | receptor | 0 | 0 | 0.644922342 |
| IFNAR2 | interferon (alpha, beta and omega) receptor 2 [Source:HGNC Symbol;Acc:HGNC:5433] | secondary receptor | 1 | 0 | 0.645534346 |
| GBP5 | guanylate binding protein 5 [Source:HGNC Symbol;Acc:HGNC:19895] | regulator | 1 | 0 | 0.645564586 |
| NCKIPSD | NCK interacting protein with SH3 domain [Source:HGNC Symbol;Acc:HGNC:15486] | signal transducer | 1 | 0 | 0.645703479 |
| NLRP8 | NLR family, pyrin domain containing 8 [Source:HGNC Symbol;Acc:HGNC:22940] | receptor | 0 | 1 | 0.646540605 |
| TLR5 | toll-like receptor 5 [Source:HGNC Symbol;Acc:HGNC:11851] | receptor | 1 | 1 | 0.646812123 |
| IFNA2 | interferon, alpha 2 [Source:HGNC Symbol;Acc:HGNC:5423] | effector | 1 | 1 | 0.646949327 |
| VNN1 | vanin 1 [Source:HGNC Symbol;Acc:HGNC:12705] | UC | 1 | 0 | 0.647088634 |
| PROS1 | protein S (alpha) [Source:HGNC Symbol;Acc:HGNC:9456] | UC | 1 | 1 | 0.649340643 |
| FCRL5 | Fc receptor-like 5 [Source:HGNC Symbol;Acc:HGNC:18508] | receptor | 0 | 0 | 0.64936884 |
| MFF | mitochondrial fission factor [Source:HGNC Symbol;Acc:HGNC:24858] | effector | 0 | 1 | 0.649515651 |
| PDGFB | platelet-derived growth factor beta polypeptide [Source:HGNC Symbol;Acc:HGNC:8800] | effector | 1 | 0 | 0.649769918 |
| KLRC2 | killer cell lectin-like receptor subfamily C, member 2 [Source:HGNC Symbol;Acc:HGNC:6375] | receptor | 1 | 0 | 0.650521543 |
| PGLYRP1 | peptidoglycan recognition protein 1 [Source:HGNC Symbol;Acc:HGNC:8904] | receptor | 1 | 1 | 0.650606353 |
| TREM2 | triggering receptor expressed on myeloid cells 2 [Source:HGNC Symbol;Acc:HGNC:17761] | receptor | 1 | 1 | 0.650926079 |
| SLX4 | SLX4 structure-specific endonuclease subunit [Source:HGNC Symbol;Acc:HGNC:23845] | regulator | 0 | 1 | 0.653314499 |
| CYBA | cytochrome b-245, alpha polypeptide [Source:HGNC Symbol;Acc:HGNC:2577] | effector | 1 | 1 | 0.653772358 |
| TREML2 | triggering receptor expressed on myeloid cells-like 2 [Source:HGNC Symbol;Acc:HGNC:21092] | receptor | 0 | 1 | 0.654322452 |
| DUOX2 | dual oxidase 2 [Source:HGNC Symbol;Acc:HGNC:13273] | accessory molecule | 0 | 1 | 0.654376699 |
| C8B | complement component 8, beta polypeptide [Source:HGNC Symbol;Acc:HGNC:1353] | signal transducer | 1 | 1 | 0.654666696 |
| NLRP12 | NLR family, pyrin domain containing 12 [Source:HGNC Symbol;Acc:HGNC:22938] | receptor | 0 | 1 | 0.656990401 |

| | | | | | |
|---|---|---|---|---|---|
| VSIG4 | V-set and immunoglobulin domain containing 4 [Source:HGNC Symbol;Acc:HGNC:17032] | receptor | 1 | 0 | 0.658144485 |
| IRAK1BP1 | interleukin-1 receptor-associated kinase 1 binding protein 1 [Source:HGNC Symbol;Acc:HGNC:17368] | signal transducer | 0 | 1 | 0.658297811 |
| CXCL2 | chemokine (C-X-C motif) ligand 2 [Source:HGNC Symbol;Acc:HGNC:4603] | effector | 0 | 1 | 0.65852338 |
| P2RX7 | purinergic receptor P2X, ligand gated ion channel, 7 [Source:HGNC Symbol;Acc:HGNC:8537] | receptor | 1 | 1 | 0.658594758 |
| BIRC2 | baculoviral IAP repeat containing 2 [Source:HGNC Symbol;Acc:HGNC:590] | regulator | 1 | 1 | 0.662579704 |
| SIGLEC8 | sialic acid binding Ig-like lectin 8 [Source:HGNC Symbol;Acc:HGNC:10877] | receptor | 0 | 1 | 0.663270305 |
| NCR2 | natural cytotoxicity triggering receptor 2 [Source:HGNC Symbol;Acc:HGNC:6732] | receptor | 1 | 0 | 0.663539057 |
| TRIM69 | tripartite motif containing 69 [Source:HGNC Symbol;Acc:HGNC:17857] | regulator | 0 | 0 | 0.66396116 |
| IL36A | interleukin 36, alpha [Source:HGNC Symbol;Acc:HGNC:15562] | effector | 1 | 0 | 0.664337557 |
| DAGLB | diacylglycerol lipase, beta [Source:HGNC Symbol;Acc:HGNC:28923] | UC | 0 | 0 | 0.665841502 |
| TMEM173 | transmembrane protein 173 [Source:HGNC Symbol;Acc:HGNC:27962] | signal transducer | 1 | 1 | 0.666766922 |
| TRIM48 | tripartite motif containing 48 [Source:HGNC Symbol;Acc:HGNC:19021] | regulator | 0 | 0 | 0.668270041 |
| TCEB2 | transcription elongation factor B (SIII), polypeptide 2 (18kDa, elongin B) [Source:HGNC Symbol;Acc:HGNC:11619] | transcription | 0 | 1 | 0.668369095 |
| S100A7 | S100 calcium binding protein A7 [Source:HGNC Symbol;Acc:HGNC:10497] | effector | 1 | 0 | 0.67326456 |
| PTX4 | pentraxin 4, long [Source:HGNC Symbol;Acc:HGNC:14171] | receptor | 0 | 0 | 0.67415034 |
| FCGR3B | Fc fragment of IgG, low affinity IIIb, receptor (CD16b) [Source:HGNC Symbol;Acc:HGNC:3620] | receptor | 0 | 0 | 0.674394711 |
| AVP | arginine vasopressin [Source:HGNC Symbol;Acc:HGNC:894] | UC | 0 | 1 | 0.674434016 |
| IL32 | interleukin 32 [Source:HGNC Symbol;Acc:HGNC:16830] | effector | 0 | 1 | 0.674874372 |
| MERTK | MER proto-oncogene, tyrosine kinase [Source:HGNC Symbol;Acc:HGNC:7027] | signal transducer | 0 | 1 | 0.67728437 |
| RAET1G | retinoic acid early transcript 1G [Source:HGNC Symbol;Acc:HGNC:16795] | effector | 1 | 0 | 0.677485993 |
| IRAK2 | interleukin-1 receptor-associated kinase 2 [Source:HGNC Symbol;Acc:HGNC:6113] | signal transducer | 1 | 1 | 0.677622456 |
| IFNA21 | interferon, alpha 21 [Source:HGNC Symbol;Acc:HGNC:5424] | effector | 1 | 0 | 0.681187556 |
| EDN1 | endothelin 1 [Source:HGNC Symbol;Acc:HGNC:3176] | UC | 1 | 1 | 0.683572779 |
| DEFB119 | defensin, beta 119 [Source:HGNC Symbol;Acc:HGNC:18099] | effector | 1 | 0 | 0.683837617 |
| DEFA3 | defensin, alpha 3, neutrophil-specific [Source:HGNC Symbol;Acc:HGNC:2762] | effector | 1 | 1 | 0.687448503 |
| IFNA16 | interferon, alpha 16 [Source:HGNC Symbol;Acc:HGNC:5421] | effector | 1 | 0 | 0.687625243 |
| SCARF1 | scavenger receptor class F, member 1 [Source:HGNC Symbol;Acc:HGNC:16820] | receptor | 0 | 1 | 0.690333526 |

| | | | | | |
|---|---|---|---|---|---|
| CASP9 | caspase 9, apoptosis-related cysteine peptidase [Source:HGNC Symbol;Acc:HGNC:1511] | signal transducer | 1 | 0 | 0.690720582 |
| C8A | complement component 8, alpha polypeptide [Source:HGNC Symbol;Acc:HGNC:1352] | signal transducer | 1 | 1 | 0.691116132 |
| CMA1 | chymase 1, mast cell [Source:HGNC Symbol;Acc:HGNC:2097] | regulator | 0 | 1 | 0.692138642 |
| CD300LF | CD300 molecule-like family member f [Source:HGNC Symbol;Acc:HGNC:29883] | regulator | 0 | 1 | 0.692278389 |
| TRIB3 | tribbles pseudokinase 3 [Source:HGNC Symbol;Acc:HGNC:16228] | regulator | 1 | 1 | 0.692648532 |
| PLA2R1 | phospholipase A2 receptor 1, 180kDa [Source:HGNC Symbol;Acc:HGNC:9042] | receptor | 0 | 0 | 0.692927233 |
| HAVCR2 | hepatitis A virus cellular receptor 2 [Source:HGNC Symbol;Acc:HGNC:18437] | receptor | 0 | 1 | 0.69357234 |
| CLEC4D | C-type lectin domain family 4, member D [Source:HGNC Symbol;Acc:HGNC:14554] | receptor | 1 | 1 | 0.694515799 |
| RNASE7 | ribonuclease, RNase A family, 7 [Source:HGNC Symbol;Acc:HGNC:19278] | effector | 1 | 0 | 0.696560696 |
| BTC | betacellulin [Source:HGNC Symbol;Acc:HGNC:1121] | effector | 1 | 0 | 0.696782055 |
| DEFB108B | defensin, beta 108B [Source:HGNC Symbol;Acc:HGNC:29966] | effector | 1 | 0 | 0.697015301 |
| PDCD1 | programmed cell death 1 [Source:HGNC Symbol;Acc:HGNC:8760] | secondary receptor | 0 | 1 | 0.698764684 |
| NOD2 | nucleotide-binding oligomerization domain containing 2 [Source:HGNC Symbol;Acc:HGNC:5331] | receptor | 1 | 1 | 0.699371876 |
| CCR3 | chemokine (C-C motif) receptor 3 [Source:HGNC Symbol;Acc:HGNC:1604] | secondary receptor | 0 | 1 | 0.702330089 |
| DCD | dermcidin [Source:HGNC Symbol;Acc:HGNC:14669] | effector | 0 | 1 | 0.703019633 |
| PLD2 | phospholipase D2 [Source:HGNC Symbol;Acc:HGNC:9068] | signal transducer | 1 | 0 | 0.704212161 |
| IFNB1 | interferon, beta 1, fibroblast [Source:HGNC Symbol;Acc:HGNC:5434] | effector | 1 | 1 | 0.704845717 |
| TRAT1 | T cell receptor associated transmembrane adaptor 1 [Source:HGNC Symbol;Acc:HGNC:30698] | regulator | 1 | 1 | 0.70639788 |
| DHX58 | DEXH (Asp-Glu-X-His) box polypeptide 58 [Source:HGNC Symbol;Acc:HGNC:29517] | receptor | 1 | 1 | 0.707786166 |
| PRTN3 | proteinase 3 [Source:HGNC Symbol;Acc:HGNC:9495] | adaptor | 0 | 1 | 0.709287208 |
| AMACR | alpha-methylacyl-CoA racemase [Source:HGNC Symbol;Acc:HGNC:451] | UC | 0 | 0 | 0.712865798 |
| CLEC4F | C-type lectin domain family 4, member F [Source:HGNC Symbol;Acc:HGNC:25357] | receptor | 0 | 0 | 0.713573522 |
| UBA7 | ubiquitin-like modifier activating enzyme 7 [Source:HGNC Symbol;Acc:HGNC:12471] | UC | 1 | 0 | 0.714235282 |
| IL6 | interleukin 6 [Source:HGNC Symbol;Acc:HGNC:6018] | effector | 0 | 1 | 0.714446015 |
| CLEC3A | C-type lectin domain family 3, member A [Source:HGNC Symbol;Acc:HGNC:2052] | receptor | 0 | 0 | 0.716309025 |
| BPIFB1 | BPI fold containing family B, member 1 [Source:HGNC Symbol;Acc:HGNC:16108] | effector | 1 | 0 | 0.72401673 |
| THEM4 | thioesterase superfamily member 4 [Source:HGNC Symbol;Acc:HGNC:17947] | regulator | 1 | 0 | 0.725070917 |

| | | | | | |
|---|---|---|---|---|---|
| FGF19 | fibroblast growth factor 19 [Source:HGNC Symbol;Acc:HGNC:3675] | effector | 1 | 0 | 0.726580768 |
| APOBEC3B | apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3B [Source:HGNC Symbol;Acc:HGNC:17352] | effector | 1 | 1 | 0.727154173 |
| UCP2 | uncoupling protein 2 (mitochondrial, proton carrier) [Source:HGNC Symbol;Acc:HGNC:12518] | UC | 0 | 1 | 0.727814267 |
| TLR4 | toll-like receptor 4 [Source:HGNC Symbol;Acc:HGNC:11850] | receptor | 1 | 1 | 0.729059965 |
| RUSC1 | RUN and SH3 domain containing 1 [Source:HGNC Symbol;Acc:HGNC:17153] | signal transducer | 0 | 1 | 0.729290463 |
| RAET1L | retinoic acid early transcript 1L [Source:HGNC Symbol;Acc:HGNC:16798] | effector | 1 | 0 | 0.733255258 |
| VENTX | VENT homeobox [Source:HGNC Symbol;Acc:HGNC:13639] | transcription | 0 | 1 | 0.733782461 |
| SLAMF9 | SLAM family member 9 [Source:HGNC Symbol;Acc:HGNC:18430] | receptor | 0 | 1 | 0.733944457 |
| CLECL1 | C-type lectin-like 1 [Source:HGNC Symbol;Acc:HGNC:24462] | receptor | 0 | 0 | 0.733959359 |
| SLC11A1 | solute carrier family 11 (proton-coupled divalent metal ion transporter), member 1 [Source:HGNC Symbol;Acc:HGNC:10907] | UC | 1 | 0 | 0.734261317 |
| MRGPRX2 | MAS-related GPR, member X2 [Source:HGNC Symbol;Acc:HGNC:17983] | receptor | 0 | 1 | 0.734291506 |
| MBL2 | mannose-binding lectin (protein C) 2, soluble [Source:HGNC Symbol;Acc:HGNC:6922] | receptor | 1 | 1 | 0.734385742 |
| AHSG | alpha-2-HS-glycoprotein [Source:HGNC Symbol;Acc:HGNC:349] | regulator | 0 | 1 | 0.734873774 |
| FCN2 | ficolin (collagen/fibrinogen domain containing lectin) 2 [Source:HGNC Symbol;Acc:HGNC:3624] | receptor | 1 | 1 | 0.736777397 |
| PDCD1LG2 | programmed cell death 1 ligand 2 [Source:HGNC Symbol;Acc:HGNC:18731] | effector | 0 | 1 | 0.736942076 |
| FCER1G | Fc fragment of IgE, high affinity I, receptor for; gamma polypeptide [Source:HGNC Symbol;Acc:HGNC:3611] | receptor | 1 | 0 | 0.737657662 |
| NRG1 | neuregulin 1 [Source:HGNC Symbol;Acc:HGNC:7997] | UC | 1 | 0 | 0.737705774 |
| C6 | complement component 6 [Source:HGNC Symbol;Acc:HGNC:1339] | signal transducer | 1 | 1 | 0.738234109 |
| UBD | ubiquitin D [Source:HGNC Symbol;Acc:HGNC:18795] | accessory molecule | 1 | 1 | 0.740072113 |
| IFNA7 | interferon, alpha 7 [Source:HGNC Symbol;Acc:HGNC:5428] | effector | 1 | 0 | 0.740935431 |
| C7 | complement component 7 [Source:HGNC Symbol;Acc:HGNC:1346] | signal transducer | 1 | 1 | 0.74679417 |
| BPI | bactericidal/permeability-increasing protein [Source:HGNC Symbol;Acc:HGNC:1095] | effector | 0 | 0 | 0.748488343 |
| TNFRSF13B | tumor necrosis factor receptor superfamily, member 13B [Source:HGNC Symbol;Acc:HGNC:18153] | secondary receptor | 0 | 1 | 0.755405322 |
| DEFB125 | defensin, beta 125 [Source:HGNC Symbol;Acc:HGNC:18105] | effector | 1 | 0 | 0.755556593 |
| NFATC3 | nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 3 [Source:HGNC | transcription | 1 | 1 | 0.758678929 |

| | | | | | |
|---|---|---|---|---|---|
| | Symbol;Acc:HGNC:7777] | | | | |
| CARD16 | caspase recruitment domain family, member 16 [Source:HGNC Symbol;Acc:HGNC:33701] | regulator | 0 | 1 | 0.759624692 |
| IFNA10 | interferon, alpha 10 [Source:HGNC Symbol;Acc:HGNC:5418] | effector | 1 | 0 | 0.760117317 |
| NLRP14 | NLR family, pyrin domain containing 14 [Source:HGNC Symbol;Acc:HGNC:22939] | receptor | 0 | 1 | 0.761356443 |
| TRIM5 | tripartite motif containing 5 [Source:HGNC Symbol;Acc:HGNC:16276] | regulator | 1 | 1 | 0.761799316 |
| RAET1E | retinoic acid early transcript 1E [Source:HGNC Symbol;Acc:HGNC:16793] | effector | 1 | 0 | 0.765811389 |
| IL1RL1 | interleukin 1 receptor-like 1 [Source:HGNC Symbol;Acc:HGNC:5998] | secondary receptor | 0 | 1 | 0.770389805 |
| NCF2 | neutrophil cytosolic factor 2 [Source:HGNC Symbol;Acc:HGNC:7661] | effector | 1 | 0 | 0.770560445 |
| REG3G | regenerating islet-derived 3 gamma [Source:HGNC Symbol;Acc:HGNC:29595] | receptor | 0 | 0 | 0.771505853 |
| ZBP1 | Z-DNA binding protein 1 [Source:HGNC Symbol;Acc:HGNC:16176] | receptor | 1 | 1 | 0.78002072 |
| VTN | vitronectin [Source:HGNC Symbol;Acc:HGNC:12724] | adaptor | 1 | 0 | 0.781454709 |
| LBP | lipopolysaccharide binding protein [Source:HGNC Symbol;Acc:HGNC:6517] | receptor | 1 | 0 | 0.785965783 |
| GBP6 | guanylate binding protein family, member 6 [Source:HGNC Symbol;Acc:HGNC:25395] | effector | 1 | 0 | 0.787882455 |
| APOBEC3H | apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3H [Source:HGNC Symbol;Acc:HGNC:24100] | effector | 1 | 0 | 0.792209211 |
| TNFRSF18 | tumor necrosis factor receptor superfamily, member 18 [Source:HGNC Symbol;Acc:HGNC:11914] | secondary receptor | 0 | 1 | 0.79293711 |
| RARRES2 | retinoic acid receptor responder (tazarotene induced) 2 [Source:HGNC Symbol;Acc:HGNC:9868] | effector | 0 | 1 | 0.795776979 |
| LILRA2 | leukocyte immunoglobulin-like receptor, subfamily A (with TM domain), member 2 [Source:HGNC Symbol;Acc:HGNC:6603] | receptor | 0 | 1 | 0.798352241 |
| CFTR | cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7) [Source:HGNC Symbol;Acc:HGNC:1884] | UC | 0 | 1 | 0.800366728 |
| ICAM1 | intercellular adhesion molecule 1 [Source:HGNC Symbol;Acc:HGNC:5344] | UC | 1 | 1 | 0.803031146 |
| DEFB1 | defensin, beta 1 [Source:HGNC Symbol;Acc:HGNC:2766] | effector | 1 | 1 | 0.808136151 |
| TLR1 | toll-like receptor 1 [Source:HGNC Symbol;Acc:HGNC:11847] | receptor | 1 | 1 | 0.819169841 |
| FCER2 | Fc fragment of IgE, low affinity II, receptor for (CD23) [Source:HGNC Symbol;Acc:HGNC:3612] | receptor | 0 | 0 | 0.822933251 |
| IFNL3 | interferon, lambda 3 [Source:HGNC Symbol;Acc:HGNC:18365] | effector | 0 | 0 | 0.847520306 |
| OAS1 | 2'-5'-oligoadenylate synthetase 1, 40/46kDa [Source:HGNC Symbol;Acc:HGNC:8086] | receptor | 1 | 1 | 0.850043663 |
| HRG | histidine-rich glycoprotein [Source:HGNC Symbol;Acc:HGNC:5181] | UC | 0 | 1 | 0.854682938 |
| DEFB132 | defensin, beta 132 [Source:HGNC Symbol;Acc:HGNC:33806] | effector | 1 | 0 | 0.858563287 |
| CARD6 | caspase recruitment domain family, member 6 [Source:HGNC Symbol;Acc:HGNC:16394] | regulator | 0 | 1 | 0.8645236 |

| | | | | | |
|---|---|---|---|---|---|
| GNLY | granulysin [Source:HGNC Symbol;Acc:HGNC:4414] | effector | 0 | 0 | 0.8963826 |
| KIR3DL2 | killer cell immunoglobulin-like receptor, three domains, long cytoplasmic tail, 2 [Source:HGNC Symbol;Acc:HGNC:6339] | receptor | 0 | 1 | 0.927148866 |
| CD36 | CD36 molecule (thrombospondin receptor) [Source:HGNC Symbol;Acc:HGNC:1663] | receptor | 1 | 1 | 0.997466675 |
| APOBEC3D | apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3D [Source:HGNC Symbol;Acc:HGNC:17354] | effector | 1 | 0 | 1.096834235 |
| APOL1 | apolipoprotein L, 1 [Source:HGNC Symbol;Acc:HGNC:618] | effector | 1 | 0 | NA |
| BRK1 | BRICK1, SCAR/WAVE actin-nucleating complex subunit [Source:HGNC Symbol;Acc:HGNC:23057] | UC | 1 | 0 | NA |
| CARD18 | caspase recruitment domain family, member 18 [Source:HGNC Symbol;Acc:HGNC:28861] | regulator | 0 | 1 | NA |
| DEFB103A | defensin, beta 103A [Source:HGNC Symbol;Acc:HGNC:15967] | effector | 1 | 1 | NA |
| DEFB104A | defensin, beta 104A [Source:HGNC Symbol;Acc:HGNC:18115] | effector | 1 | 0 | NA |
| DEFB105A | defensin, beta 105A [Source:HGNC Symbol;Acc:HGNC:18087] | effector | 0 | 0 | NA |
| DEFB106A | defensin, beta 106A [Source:HGNC Symbol;Acc:HGNC:18088] | effector | 1 | 0 | NA |
| DEFB107A | defensin, beta 107A [Source:HGNC Symbol;Acc:HGNC:18086] | effector | 0 | 0 | NA |
| DEFB108P1 | defensin, beta 108, pseudogene 1 [Source:HGNC Symbol;Acc:HGNC:18089] | effector | 0 | 0 | NA |
| DEFB108P2 | defensin, beta 108, pseudogene 2 [Source:HGNC Symbol;Acc:HGNC:30847] | effector | 0 | 0 | NA |
| DEFB109P1 | defensin, beta 109, pseudogene 1 [Source:HGNC Symbol;Acc:HGNC:18090] | effector | 0 | 0 | NA |
| DEFB109P1B | defensin, beta 109, pseudogene 1B [Source:HGNC Symbol;Acc:HGNC:33469] | effector | 0 | 0 | NA |
| DEFB128 | defensin, beta 128 [Source:HGNC Symbol;Acc:HGNC:18106] | effector | 1 | 0 | NA |
| DEFB4A | defensin, beta 4A [Source:HGNC Symbol;Acc:HGNC:2767] | effector | 1 | 0 | NA |
| EGLN2 | egl-9 family hypoxia-inducible factor 2 [Source:HGNC Symbol;Acc:HGNC:14660] | UC | 0 | 1 | NA |
| FCGR1A | Fc fragment of IgG, high affinity Ia, receptor (CD64) [Source:HGNC Symbol;Acc:HGNC:3613] | receptor | 1 | 1 | NA |
| FCGR1B | Fc fragment of IgG, high affinity Ib, receptor (CD64) [Source:HGNC Symbol;Acc:HGNC:3614] | receptor | 1 | 0 | NA |
| FRK | fyn-related Src family tyrosine kinase [Source:HGNC Symbol;Acc:HGNC:3955] | regulator | 1 | 0 | NA |
| GPER1 | G protein-coupled estrogen receptor 1 [Source:HGNC Symbol;Acc:HGNC:4485] | secondary receptor | 1 | 0 | NA |
| GPR33 | G protein-coupled receptor 33 (gene/pseudogene) [Source:HGNC Symbol;Acc:HGNC:4489] | secondary receptor | 0 | 1 | NA |
| H2BFS | H2B histone family, member S (pseudogene) [Source:HGNC Symbol;Acc:HGNC:4762] | UC | 1 | 0 | NA |
| HMGN2 | high mobility group nucleosomal binding domain 2 [Source:HGNC Symbol;Acc:HGNC:4986] | regulator | 0 | 1 | NA |
| IL36B | interleukin 36, beta [Source:HGNC Symbol;Acc:HGNC:15564] | effector | 1 | 0 | NA |
| IL37 | interleukin 37 [Source:HGNC Symbol;Acc:HGNC:15563] | effector | 0 | 0 | NA |

| | | | | | |
|---|---|---|---|---|---|
| CXCL8 | chemokine (C-X-C motif) ligand 8 [Source:HGNC Symbol;Acc:HGNC:6025] | effector | 0 | 0 | NA |
| IRF3 | interferon regulatory factor 3 [Source:HGNC Symbol;Acc:HGNC:6118] | transcription | 1 | 1 | NA |
| KIR2DS1 | killer cell immunoglobulin-like receptor, two domains, short cytoplasmic tail, 1 [Source:HGNC Symbol;Acc:HGNC:6333] | receptor | 1 | 0 | NA |
| KIR2DS2 | killer cell immunoglobulin-like receptor, two domains, short cytoplasmic tail, 2 [Source:HGNC Symbol;Acc:HGNC:6334] | receptor | 1 | 0 | NA |
| KIR2DS4 | killer cell immunoglobulin-like receptor, two domains, short cytoplasmic tail, 4 [Source:HGNC Symbol;Acc:HGNC:6336] | receptor | 1 | 0 | NA |
| KIR2DS5 | killer cell immunoglobulin-like receptor, two domains, short cytoplasmic tail, 5 [Source:HGNC Symbol;Acc:HGNC:6337] | receptor | 1 | 0 | NA |
| KIR3DS1 | killer cell immunoglobulin-like receptor, three domains, short cytoplasmic tail, 1 [Source:HGNC Symbol;Acc:HGNC:6340] | receptor | 1 | 0 | NA |
| MAP3K14 | mitogen-activated protein kinase kinase kinase 14 [Source:HGNC Symbol;Acc:HGNC:6853] | signal transducer | 0 | 1 | NA |
| MMP12 | matrix metallopeptidase 12 (macrophage elastase) [Source:HGNC Symbol;Acc:HGNC:7158] | effector | 0 | 1 | NA |
| MRC1 | mannose receptor, C type 1 [Source:HGNC Symbol;Acc:HGNC:7228] | receptor | 1 | 0 | NA |
| NCAM1 | neural cell adhesion molecule 1 [Source:HGNC Symbol;Acc:HGNC:7656] | UC | 1 | 0 | NA |
| NCF1 | neutrophil cytosolic factor 1 [Source:HGNC Symbol;Acc:HGNC:7660] | effector | 1 | 0 | NA |
| NLRC3 | NLR family, CARD domain containing 3 [Source:HGNC Symbol;Acc:HGNC:29889] | signal transducer | 0 | 1 | NA |
| OTULIN | OTU deubiquitinase with linear linkage specificity [Source:HGNC Symbol;Acc:HGNC:25118] | regulator | 1 | 0 | NA |
| PPP3R1 | protein phosphatase 3, regulatory subunit B, alpha [Source:HGNC Symbol;Acc:HGNC:9317] | signal transducer | 1 | 1 | NA |
| PRSS2 | NA | effector | 1 | 0 | NA |
| PSTPIP1 | proline-serine-threonine phosphatase interacting protein 1 [Source:HGNC Symbol;Acc:HGNC:9580] | adaptor | 1 | 1 | NA |
| RNASE2 | ribonuclease, RNase A family, 2 (liver, eosinophil-derived neurotoxin) [Source:HGNC Symbol;Acc:HGNC:10045] | effector | 0 | 0 | NA |
| RNASE3 | ribonuclease, RNase A family, 3 [Source:HGNC Symbol;Acc:HGNC:10046] | effector | 1 | 0 | NA |
| RPL13A | ribosomal protein L13a [Source:HGNC Symbol;Acc:HGNC:10304] | signal transducer | 1 | 0 | NA |
| RPS19 | ribosomal protein S19 [Source:HGNC Symbol;Acc:HGNC:10402] | accessory molecule | 0 | 1 | NA |
| SDHAF4 | NA | accessory molecule | 1 | 0 | NA |

| | | | | | |
|---|---|---|---|---|---|
| SIGLEC16 | sialic acid binding Ig-like lectin 16 (gene/pseudogene) [Source:HGNC Symbol;Acc:HGNC:24851] | receptor | 1 | 0 | NA |
| SRXN1 | sulfiredoxin 1 [Source:HGNC Symbol;Acc:HGNC:16132] | UC | 0 | 1 | NA |
| STAP2 | signal transducing adaptor family member 2 [Source:HGNC Symbol;Acc:HGNC:30430] | regulator | 0 | 1 | NA |
| TPSB2 | tryptase beta 2 (gene/pseudogene) [Source:HGNC Symbol;Acc:HGNC:14120] | effector | 0 | 1 | NA |
| TRIL | TLR4 interactor with leucine-rich repeats [Source:HGNC Symbol;Acc:HGNC:22200] | signal transducer | 1 | 0 | NA |
| TRIM53AP | tripartite motif containing 53A, pseudogene [Source:HGNC Symbol;Acc:HGNC:19025] | regulator | 0 | 0 | NA |
| TRIM64 | tripartite motif containing 64 [Source:HGNC Symbol;Acc:HGNC:14663] | regulator | 0 | 0 | NA |
| TRIM75P | tripartite motif containing 75, pseudogene [Source:HGNC Symbol;Acc:HGNC:32686] | regulator | 0 | 0 | NA |
| UBA52 | ubiquitin A-52 residue ribosomal protein fusion product 1 [Source:HGNC Symbol;Acc:HGNC:12458] | accessory molecule | 1 | 0 | NA |
| UBE2N | ubiquitin-conjugating enzyme E2N [Source:HGNC Symbol;Acc:HGNC:12492] | accessory molecule | 1 | 1 | NA |
| ULBP1 | UL16 binding protein 1 [Source:HGNC Symbol;Acc:HGNC:14893] | effector | 1 | 0 | NA |
| VASP | vasodilator-stimulated phosphoprotein [Source:HGNC Symbol;Acc:HGNC:12652] | signal transducer | 0 | 1 | NA |
| VAV1 | vav 1 guanine nucleotide exchange factor [Source:HGNC Symbol;Acc:HGNC:12657] | signal transducer | 1 | 0 | NA |
| VTRNA2-1 | vault RNA 2-1 [Source:HGNC Symbol;Acc:HGNC:37054] | UC | 0 | 1 | NA |
| ZFP36 | ZFP36 ring finger protein [Source:HGNC Symbol;Acc:HGNC:12862] | effector | 0 | 0 | NA |

**Table S2.** Enrichment analyses of SNPs presenting extreme (top 1%) $F_{CS}$ values

| Population | Genic vs. nongenic[a] | IIGs vs. remainder [b] |
|---|---|---|
| YRI | 1.276*** | 1.271 |
| CEU | 1.266*** | 0.910 |
| CHB | 1.215*** | 1.210 |

[a]Enrichment of genic SNPs with respect to nongenic variants
[b]Enrichment of genic SNPs located in innate immunity genes with respect to non-innate immunity genes and nongenic variants (remainder)

***$P$-value $<10^{-4}$

**Table S3.** Innate immunity genes identified as candidates for positive selection and reported as associated with humain traits by genome-wide association studies (GWAs)

| Candidate gene | Population[a] | Trait |
|---|---|---|
| *ADAM15* | ASN | Prostate cancer |
| *ARHGEF7* | ASN | Menopause |
| *ARPC1A* | CEU | Dehydroepiandrosterone sulphate levels |
| *BLK* | ASN | Rheumatoid arthritis, Systemic lupus erythematosus, Kawasaki disease |
| *CCDC88A* | YRI, CEU | Height |
| *CCR2* | YRI | Celiac disease, Obesity-related traits, Celiac disease |
| *CD36* | YRI | HDL cholesterol, Mean platelet volume |
| *CD80* | YRI | Systemic lupus erythematosus, Primary biliary cirrhosis, Vitiligo, Celiac disease |
| *CHUK* | YRI | Liver enzyme levels |
| *ERBB4* | ASN | Breast cancer |
| *EYA4* | CEU | Bone mineral density |
| *FCER1A* | YRI | Obesity-related traits, C-reactive protein, IgE levels |
| *FER* | ASN | Adiponectin levels, Height |
| *IFIH1* | YRI | Inflammatory bowel disease, Vitiligo, Type 1 diabetes, Type 1 diabetes autoantibodies, Psoriasis, Immunoglobulin A |
| *IL4* | CEU | Inflammatory bowel disease |
| *IRGM* | CEU | Inflammatory bowel disease, Crohn's disease |
| *ITPR3* | ASN | Crohn's disease, Phosphorus levels |
| *LTBP1* | YRI | Height, Immune response to smallpox vaccine |
| *MAP3K1* | CEU | Breast cancer, Triglycerides |
| *MAPK10* | YRI | Liver enzyme levels |
| *MERTK* | ASN | Hepatitis C induced liver fibrosis, Multiple sclerosis |
| *MYH9* | YRI | Glomerulosclerosis, End-stage renal disease |
| *NRG1* | ASN | Thyroid hormone levels, Thyroid cancer, Hirschsprung's disease |
| *RAF1* | ASN | Total cholesterol |
| *TLR1* | CEU | Asthma and hay fever, Allergic sensitization, Self-reported allergy, *Helicobacter pylori* serologic status |
| *TNRC6B* | YRI | Uterine fibroids |

| | | |
|---|---|---|
| *ZFPM2* | YRI | Glaucoma, Platelet counts, Vascular endothelial growth factor levels |

[a]Population where the FCS signal was the strongest

**Table S4.** SNPs identified in genome-wide association studies (GWAs) that are outliers, or in LD ($r^2$>0.8) with outliers, identified in our genome-wide scan for positive selection

| Candidate SNP[a] | Chromosome | Position | Population[b] | %FCS[c] | Region max %FCS | DAF[d] | Candidate gene | Closest gene | GWAs SNP[e] | Phenotype |
|---|---|---|---|---|---|---|---|---|---|---|
| rs10796955 | 1 | 155959303 | CHB | 6.0% | 12.1% | 0.856 | *ARHGEF2* | -- | rs670523 ($r^2$=0.91) | Inflammatory bowel disease |
| rs2511215 | 1 | 159255749 | YRI | -- | 9.8% | 0.023 | *FCER1A* | -- | rs2251746 ($r^2$=1) | IgE levels |
| rs2589113 | 2 | 555163023 | YRI | 4.3% | 4.4% | 0.403 | *CCDC88A* | *CCDC88A* | rs2589113 | Height |
| rs6709662 | 2 | 112772465 | CHB | 10.1% | 11.0% | 0.789 | *MERTK* | *MERTK* | rs4374383 ($r^2$=1) | Hepatitis C induced liver fibrosis |
| rs7648488 | 3 | 127089046 | CHB | 9.7% | 16.6% | 0.964 | *RAF1* | -- | rs2290159 ($r^2$=1) | Total cholesterol |
| rs5863951 | 3 | 119199157 | YRI | 4.7% | 4.9% | 0.443 | *CD80* | *KTELC1* | rs16829840 ($r^2$=1) | Plasma omega-6 polyunsaturated fatty acid levels |
| rs10004195 | 4 | 387847024 | CEU | 6.9% | 9.4% | 0.818 | *TLR1* | -- | rs10004195 | *Helicobacter pylori* serologic status |
| rs4833095 | 4 | 387997010 | CEU | 7.5% | 9.4% | 0.824 | *TLR1* | *TLR1* | rs4833095 | Asthma and hay fever |
| rs2101521 | 4 | 388115051 | CEU | 8.5% | 9.4% | 0.835 | *TLR1* | -- | rs2101521 | Self-reported allergy |
| rs1761643 | 4 | 388128076 | CEU | 8.6% | 9.4% | 0.824 | *TLR1* | -- | rs17616434 | Allergic sensitization |
| rs1295686 | 5 | 131995843 | CEU | 5.5% | 13.2% | 0.767 | *IL4* | *IL13* | rs1295686 | Atopic dermatitis |
| rs848 | 5 | 131996500 | CEU | 5.8% | 13.2% | 0.820 | *IL4* | *IL13* | rs20541 ($r^2$=0.97) | Hodgkin's lymphoma; IgE levels; Psoriasis |
| rs2897442 | 5 | 132049027 | CEU | 9.6% | 13.2% | 0.820 | *IL4* | *KIF3A* | rs2897442 | Atopic dermatitis |
| rs13361189 | 5 | 150223387 | CEU | 6.9% | 10.0% | 0.941 | *IRGM* | -- | rs13361189 | Crohn's disease |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| rs1005180 | | 150226 | | 7.0 | | 0.9 | | | rs11747270 | Inflammatory bowel disease, Crohn's |
| 4 | 5 | 424 | CEU | % | 10.0% | 41 | *IRGM* | *IRGM* | ($r^2$=1) | disease |
| | | 150270 | | | | 0.0 | | | | |
| rs7714584 | 5 | 420 | CEU | -- | 10.0% | 59 | *IRGM* | -- | rs7714584 | Crohn's disease |
| | | 336590 | | 8.5 | | 0.8 | | | | |
| rs4711336 | 6 | 46 | CHB | % | 10.3% | 92 | *ITPR3* | *ITPR3* | rs4711336 | Height |
| | | 336951 | | 10. | | 0.7 | | | rs751728 | |
| rs2281830 | 6 | 30 | CHB | 0% | 10.3% | 37 | *ITPR3* | *IP6K3* | ($r^2$=0.84) | Crohn's disease |
| | | 803406 | | 6.3 | | 0.3 | | | | |
| rs2366858 | 7 | 22 | YRI | % | 7.1% | 52 | *CD36* | -- | rs2366858 | HDL cholesterol |
| rs6248906 | | 113383 | | 1.0 | | 0.2 | | | rs7812879 | |
| 9 | 8 | 83 | CHB | % | 12.7% | 27 | *BLK* | -- | ($r^2$=0.97) | Systemic lupus erythematosus |
| rs1687955 | | 324112 | | 7.3 | | 0.6 | | | | |
| 2 | 8 | 16 | CHB | % | 10.9% | 24 | *NRG1* | *NRG1* | rs16879552 | Hirschsprung's disease |

[a]SNP presenting an individual FCS among the 1% highest of the genome. When several candidate SNPs were in similar LD levels with the GWAs SNP, the candidate SNP with the highest FCS was reported.
[b]Population where the FCS signal was the strongest
[c]Proportion of outlier FCS values within a 100kb window centered on the candidate SNP
[d]Derived allele frequency
[e]LD measure is reported in brackets in the relevant population when candidate SNP and GWAs SNP are not the same, but are in strong LD. GWAs SNPs are sometimes not outliers because they present a highly frequent ancestral allele, contrarily to the candidate SNP.

**Table S5**. Accuracy indices of the ABC estimations of the age of selection

| Population | ABC method | Statistics | rEB[a] | 95%COV[b] | PE[c] |
|---|---|---|---|---|---|
| **YRI** | *ridge*[d] | set 1[e] | 0.66 | 0.92 | 0.50 |
| | | set 2[f] | 0.52 | 0.92 | 0.39 |
| | *neuralnet*[d] | set 1 | 0.64 | 0.93 | 0.47 |
| | | set 2 | 0.50 | 0.95 | 0.30 |
| | *boosted-neuralnet*[g] | set 3[g] | 0.24 | 0.94 | 0.12 |
| **CEU** | ridge | set 1 | 0.66 | 0.96 | 0.55 |
| | | set 2 | 0.56 | 0.95 | 0.5 |
| | *neuralnet* | set 1 | 0.60 | 0.96 | 0.48 |
| | | set 2 | 0.33 | 0.97 | 0.26 |
| | *boosted-neuralnet* | set 3 | 0.09 | 0.97 | 0.06 |
| **CHB** | ridge | set 1 | 0.47 | 0.94 | 0.59 |
| | | set 2 | 0.38 | 0.93 | 0.46 |
| | *neuralnet* | set 1 | 0.42 | 0.94 | 0.49 |
| | | set 2 | 0.25 | 0.95 | 0.23 |
| | *boosted-neuralnet* | set 3 | 0.11 | 0.94 | 0. 10 |

[a] rEB is the relative estimation bias.
[b] 95%COV is the coverage of the 95% credible interval, i.e., the percent of times where true values were found within the 95% credible intervals obtained from posterior distributions.
[c] PE is the prediction error, i.e., the mean square error, MSE, divided by the prior variance of the parameter.
[d] ABC methods implemented in the 'abc' R package. To avoid estimations exceeding the prior bounds, we systematically used a log transformation of *t*, which has no effect on the accuracy of estimations.
[e] set 1 of summary statistics: $\boldsymbol{\Theta}_S$, $\boldsymbol{\Theta}_\pi$, Tajima's *D*, Fay and Wu *H*, iHS, DIND and $F_{ST}$ computed for the selected SNPs, using a window of 100 kb around the mutation (see Material and Methods).
[f] set 2 of summary statistics: the set 1 with their corresponding averages and proportions of 1% top values computed over the 100 kb region surrounding the selected SNP.
[g] To boost the 'neuralnet' ABC estimations, we applied a transformation of the summary statistics of the set 2. We used the transformation $T(S_i S_{j\geq i})$ with $S_i$ and $S_j$ the i$^{th}$ and j$^{th}$ summary statistics in the set 2. The set 3 of summary statistics thus corresponds to the set 2 with these transformations.

**Table S6**. ABC estimation of the age and intensity of selection using simulated data

| | Simulated parameter[a] | $\widehat{\boldsymbol{\theta}}$[b] | *95%*CI [b] | |
|---|---|---|---|---|
| **Age of selection (t)** | 5,275 | 5,325 | 2,400 | 8,825 |
| | 16,600 | 20,725 | 9,600 | 32,625 |
| | 36,575 | 40,875 | 24,000 | 58,400 |
| **Intensity of selection (s)** | 0.034 | 0.0354 | 0.0256 | 0.0447 |
| | 0.0175 | 0.0152 | 0.0077 | 0.0237 |
| | 0.0104 | 0.0105 | 0.005 | 0.0169 |

[a] True value of the simulated parameters *t* and *s* (age and intensity of selection, Material and Methods). *t* is given in years.
[b] ABC punctual estimate (posterior mean) of *t* and *s*.
[c] 95%CI is the 95% credible interval computed from the posterior distribution (Material and Methods).

**Table S7**. ABC estimations of the age of selection of the high-scoring coding variants

| Population | Gene | Chromosome | SNP | $\hat{t}$[a] | *95%*CI[b] | |
|---|---|---|---|---|---|---|
| CEU | *LCT*[c] | 2 | rs4988235 | 7,075 | 3,500 | 10,975 |
| YRI | *IFIH1* | 2 | rs10930046 | 5,425 | 3,000 | 7,900 |
| YRI | *LTF* | 3 | rs60938611 | 13,475 | 6,675 | 20,450 |
| YRI | *CD36* | 7 | rs3211938 | 3,600 | 2,125 | 5,025 |
| YRI | *ZFPM2* | 8 | rs11993776 | 13,150 | 7,550 | 18,400 |
| YRI | *DAK* | 11 | rs2260655 | 14,800 | 8,750 | 20,700 |
| CEU | *TLR1* | 4 | rs5743618 | 11,350 | 6,425 | 16,450 |
| CEU | *TLR1* | 4 | rs4833095 | 11,200 | 6,425 | 16,500 |
| CEU | *TLR6* | 4 | rs5743810 | 10,600 | 5,525 | 16,250 |
| CEU | *MAP3K1* | 5 | rs702689 | 6,700 | 3,600 | 9,925 |
| CHB | *MERTK* | 2 | rs7604639 | 9,500 | 5,975 | 13,250 |
| CHB | *MERTK* | 2 | rs2230515 | 6,775 | 4,175 | 9,700 |
| CHB | *CLEC3B* | 3 | rs13963 | 35,500 | 23,125 | 47,550 |
| CHB | *NRG1* | 8 | rs3924999 | 3,875 | 2,075 | 5,750 |

[a] $\hat{t}$ ABC posterior estimate (posterior mean) of the age of selection.
[b] 95%CI is the 95% credible interval computed for the posterior distribution of *t*.
[c] The age of selection of the C/T-13910 polymorphism (rs4988235) in the *LCT* region, known to be under strong selection in Europeans, is given for comparison

**Table S8. Innate immunity genes belonging to 5% of genes with the highest Neanderthal introgression scores at the genome-wide level**

##Introgression score: average of the probabilities of Neanderthal ancetry at all bases of the gene

| **EUROPEANS** | | | **ASIANS** | | |
|---|---|---|---|---|---|
| ***Gene name*** | ***Introgression score*** | ***Innate immune gene category*** | ***Gene name*** | ***Introgression score*** | ***Innate immune gene category*** |
| IL17F | 0.413177083 | effector | SRMS | 0.620734694 | regulator |
| CLEC3A | 0.366774359 | receptor | TXN | 0.606642857 | regulator |
| OAS3 | 0.345878345 | receptor | GNAI2 | 0.595266272 | regulator |
| OAS1 | 0.340748252 | receptor | PTK6 | 0.570146341 | regulator |
| OAS2 | 0.338545455 | receptor | TLR1 | 0.485242038 | receptor |
| SRMS | 0.334061224 | regulator | TLR10 | 0.452341969 | receptor |
| IL17A | 0.318396226 | effector | IRF6 | 0.412375566 | transcription |
| PTK6 | 0.299292683 | regulator | TLR6 | 0.370157058 | receptor |
| IFITM2 | 0.258619048 | effector | CTSH | 0.354414573 | accessory molecule |
| IFITM1 | 0.258 | effector | NLRC5 | 0.342486142 | receptor |
| IFITM3 | 0.249866667 | effector | SCARB1 | 0.340875598 | receptor |
| PRKCQ | 0.24707373 | signal transducer | SIRT1 | 0.330384236 | UC |
| IL18 | 0.222517857 | effector | MMP7 | 0.327130137 | accessory molecule |
| MMP7 | 0.221116438 | accessory molecule | PDCD1 | 0.322775 | secondary receptor |
| ACTR2 | 0.213353153 | UC | ATF3 | 0.311124654 | transcription |
| AVP | 0.208 | UC | HSP90B1 | 0.305778325 | accessory molecule |
| ARRB2 | 0.206485507 | regulator | BCL2A1 | 0.295177083 | regulator |
| ITPR3 | 0.198685777 | UC | IL17F | 0.28203125 | effector |
| ITPR2 | 0.195709861 | UC | IL17A | 0.281603774 | effector |
| TLR1 | 0.192923567 | receptor | TRIM32 | 0.267069767 | regulator |
| EPRS | 0.191293037 | regulator | IGF1 | 0.265382953 | effector |
| NLRP6 | 0.190283951 | receptor | NRIP1 | 0.255464052 | transcription |
| MAVS | 0.18945481 | adaptor | MAPK1 | 0.248505882 | signal transducer |
| FCGR2A | 0.189 | receptor | GJA1 | 0.236630435 | UC |
| TLR10 | 0.188072539 | receptor | EGF | 0.232630335 | effector |
| TRIM62 | 0.183079625 | regulator | MAP3K1 | 0.230381454 | signal transducer |
| RNF19B | 0.182762048 | effector | CDKN2A | 0.220972881 | regulator |
| KRT1 | 0.18 | regulator | FOS | 0.219970588 | transcription |
| FCGR3A | 0.177138889 | receptor | SLC15A4 | 0.217471933 | accessory molecule |
| FCGR3B | 0.177 | receptor | DUSP16 | 0.214877295 | regulator |
| KAT2B | 0.171095402 | UC | RNF19B | 0.209710843 | effector |
| TRIML1 | 0.16913913 | regulator | NFKBIE | 0.206949367 | signal transducer |

| | | | | | |
|---|---|---|---|---|---|
| KRAS | 0.16452509 | signal transducer | OAS1 | 0.205195804 | receptor |
| FCGR2B | 0.162988636 | receptor | HSP90AB1 | 0.200965217 | accessory molecule |
| TLR2 | 0.162065972 | receptor | UBC | 0.20025 | accessory molecule |
| DDX21 | 0.161697279 | adaptor | APAF1 | 0.196444995 | signal transducer |
| MMP12 | 0.159 | effector | KRT1 | 0.195276923 | regulator |
| TLR6 | 0.156037773 | receptor | CLEC3A | 0.194953846 | receptor |
| CLEC11A | 0.15275 | receptor | SYK | 0.193122024 | adaptor |
| BIRC5 | 0.147313253 | regulator | ABI1 | 0.190592347 | UC |
| SLAMF8 | 0.146306452 | receptor | CLEC2L | 0.18949434 | receptor |
| MAP1LC3C | 0.144177778 | effector | GRK5 | 0.178014661 | regulator |
| PRF1 | 0.133326531 | effector | AVP | 0.1771875 | UC |
| PLA2R1 | 0.131436316 | receptor | NXN | 0.176891468 | signal transducer |
| NOXA1 | 0.130979452 | effector | CAMK2D | 0.176760812 | signal transducer |
| PRKCG | 0.128846667 | signal transducer | MAP1LC3C | 0.175266667 | effector |
| ACAP1 | 0.12849375 | UC | TP73 | 0.174627803 | transcription |
| ELANE | 0.128052632 | regulator | TRIM62 | 0.173196721 | regulator |
| LY75 | 0.128027929 | receptor | CCR4 | 0.172444444 | secondary receptor |
| TNFSF9 | 0.12772 | effector | OAS3 | 0.17092944 | receptor |
| NFATC1 | 0.127659794 | transcription | TRIM71 | 0.164426009 | regulator |
| FOS | 0.127323529 | transcription | MT2A | 0.164307692 | UC |
| PRTN3 | 0.123526316 | adaptor | EYA4 | 0.162470397 | regulator |
| HSP90AB1 | 0.121756522 | accessory molecule | IFNA1 | 0.161 | effector |
| TRIB3 | 0.120890411 | regulator | IFNA8 | 0.159 | effector |
| IRF2 | 0.118775424 | transcription | IFNE | 0.157294118 | effector |
| CD80 | 0.117950893 | effector | POLR2E | 0.157140127 | receptor |
| PPARGC1B | 0.11721028 | signal transducer | BDKRB2 | 0.155390048 | receptor |
| ATF3 | 0.117127424 | transcription | TSC2 | 0.155374134 | regulator |
| AQP3 | 0.116892308 | UC | DMBT1 | 0.155037736 | effector |
| CLEC2L | 0.116215094 | receptor | TDGF1 | 0.151357143 | UC |
| ERAP1 | 0.115880419 | UC | CD8A | 0.148683398 | UC |
| CXCL16 | 0.115569892 | effector | ATG5 | 0.139259566 | regulator |
| TXN | 0.114541667 | regulator | TRIM29 | 0.137452991 | regulator |
| CD209 | 0.114383117 | receptor | SOS1 | 0.137398402 | regulator |
| TNK1 | 0.114104839 | signal transducer | BCL3 | 0.13727027 | signal transducer |
| CLEC4G | 0.114 | receptor | CFI | 0.136709503 | regulator |
| TNFRSF13B | 0.111490446 | secondary receptor | PRKCG | 0.135956667 | signal transducer |
| HSP90B1 | 0.107916256 | accessory molecule | ACAP1 | 0.13381875 | UC |
| ART1 | 0.10789823 | UC | IFITM1 | 0.133 | effector |
| NFKBIE | 0.107835443 | signal transducer | IFNA2 | 0.132 | effector |
| SLC15A4 | 0.106698545 | accessory | IFITM2 | 0.130666667 | effector |

| | | | | | |
|---|---|---|---|---|---|
| | | molecule | | | |
| BTN3A3 | 0.10541358 | receptor | SMAD6 | 0.12898045 | signal transducer |
| CD302 | 0.105 | receptor | CD22 | 0.126774306 | regulator |
| CAMKK2 | 0.104469974 | signal transducer | PPP3R1 | 0.126577065 | signal transducer |
| JAK3 | 0.102263699 | signal transducer | REG3G | 0.126 | receptor |

# 3. Complément d'informations pour l'article 2

# Supplemental Material

**A Genomic Portrait of the Genetic Architecture and Regulatory Impact of microRNA Expression in Response to Infection**

Katherine J. Siddle, Matthieu Deschamps, Ludovic Tailleux, Yohann Nédélec, Julien Pothlichet, Geanncarlo Lugo-Villarino, Valentina Libri, Brigitte Gicquel, Olivier Neyrolles, Guillaume Laval, Etienne Patin, Luis B. Barreiro, Lluís Quintana-Murci

**Supplemental Methods**

**Supplemental Figures**



**Additional files provided as separate files.**

Tables listed below are provided as one Excel file with multiple worksheets.

## Supplemental Methods

### miRNA expression analyses

Genome-wide miRNA expression was profiled using the Agilent Human miRNA microarray (Release 16.0) consisting of 56,044 probes representing 1205 human and 144 human viral miRNAs annotated in Build 16 of miRBase (http://www.mirbase.org/) (Griffiths-Jones et al. 2006). To minimize potential batch effects, samples were randomized across array batches while ensuring that infected and non-infected samples from the same individual were kept in the same batch. To confirm the technical reproducibility of our arrays, we performed technical replicates for 11 samples. We found miRNA expression to be highly correlated between these replicates (mean Pearson's r = 0.96, compared to r=0.93 between individuals in the same condition). Unless otherwise indicated, all analyses were carried out using the R statistical framework.

Initial analysis and quality control of the microarrays were performed using Agilent's Feature Extraction Software. Subsequent pre-processing was performed using the Bioconductor package AgiMicroRna (Lopez-Romero 2011). All arrays were normalized using the Robust Multi-Array Average (RMA) method including background correction (Irizarry et al. 2006). Probes for which expression was not detected, or was indistinguishable from background levels in at least 10% of samples (N<7) in both the non-infected and infected state, were removed. Using Principal Component (PC) analysis, we identified two samples as outliers, which were discarded from the analyses, owing to low number of expressed miRNAs. We further corrected for batch effects and removed the first PC, using a linear model, as this was associated with technical variation in sample processing.

To identify differentially expressed miRNAs upon MTB infection, we applied a linear model with a fixed effect for MTB treatment. Moderated statistics were obtained using the empirical Bayes approach, implemented in the Bioconductor package limma, (Smyth 2004) and multiple-testing corrected p-values were calculated using the Benjamini and Hochberg FDR (Benjamini and Hochberg 1995).

### Mapping of expression quantitative trait loci (eQTLs)

Associations between SNP genotypes (GEO Accession Number GSE34588) (Barreiro et al. 2012) and miRNA expression levels were calculated using a linear regression model, assuming an additive effect of alleles on expression, in infected and non-infected samples. Given the reduced power to detect associations for rare SNPs and infrequently expressed miRNAs, we restricted eQTL mapping to SNPs with a minor allele frequency higher than 10% (N=570,803) and miRNAs that were

expressed in at least 50% of samples in a given condition (N=266; 250 and 264 in non-infected and infected samples, respectively). We improved the power to detect eQTLs by quantile normalization and through the regression of a given number of PCs (1 and 4 for non-infected and infected samples, respectively), to account for unknown confounders (Supplemental Fig. S9A,B). The number of PCs removed was determined based on maximizing the number of significant associations in a given condition. However, correlations between p-values with and without PC removal were high (>0.7), showing a negligible impact on the relationship between genotypes and expression levels (Supplemental Fig. S9C).

We mapped putative cis-eQTLs using a region of 200 kb centered on the mature miRNA and recorded the lowest p-value obtained by regressing the expression level of each miRNA against the genotype of each SNP within the 200 kb window. We removed all probes that mapped to more than one genomic location. We estimated the FDR by comparing the observed to the null distribution, generated using the lowest p-values observed for each miRNA in 100 permutations of expression values (Pickrell et al. 2010; Barreiro et al. 2012). We detected genotype-treatment interaction effects by Bayesian regression with the software BRIdGE (Maranville et al. 2011), using scripts available on the authors' web pages. We used default effect sizes (0.8, 1, 1.2 & 1.6), a threshold of 0.001, and a posterior probability cut-off of 0.7 for determining significance. We mapped trans-eQTLs by performing a genome-wide association of miRNA expression levels against all genotyped SNPs, as well as against a subset of SNPs (N=4) previously identified as susceptibility loci for TB by GWAs (http://www.genome.gov/26525384) (Hindorff et al. 2009). Multiple testing corrections were performed using a Bonferroni correction at the 95% significance level.

To study the genomic context of these miR-eQTLs, we assessed their overlap with active genomic regions, using ChIP-seq and DNase-seq peak data for human monocytes (RO1746) from the ENCODE project (http://encodeproject.org/ENCODE/) (The ENCODE Project Consortium 2012). We used a Fisher's exact test to calculate enrichments of miR-eQTLs in regions associated with histone modifications or open chromatin.

For the fine-mapping of miR-eQTL regions, we imputed genotypes for SNPs not present on our genotyping array with IMPUTE2 (Howie et al. 2009), using integrated haplotype data from Phase 1 of the 1000 Genomes project (The 1000 Genomes Project Consortium 2012). We defined sets of SNPs in high linkage disequilibrium (LD, $r^2 > 0.8$) with our array-based eQTL SNPs, to test for their presence among dsQTLs (Degner et al. 2012) and mRNA-eQTLs (Barreiro et al. 2012; Xia et al. 2012). To refine miR-eQTL signals, we repeated eQTL mapping, as described above, using all array-based and imputed SNPs within a 1 Mb region around the initial set of detected miR-eQTLs. Regional associations were visualized using LocusZoom (Pruim et al. 2010).

As the power to detect a significant association between a genotype and the expression of a given gene depends on a combination of factors, including; sample size, minor allele frequency (MAF), expression level, inter-individual variability in expression and the fold change between genotypes, we performed simulations to quantify the power of the present study to map eQTLs varying these parameters (Supplemental Fig. S2). For sample sizes ranging from 20-150 individuals we first simulated genotypes for a SNP with a MAF of either 0.2 or 0.5. We next simulated expression levels for each genotype using a normal distribution with mean and standard deviation derived from our observed data, and varying the fold change between the mean expression levels of the homozygote genotypes. Simulated expression levels were then quantile normalized, across all genotypes, and the association calculated using a linear model as described above. We performed 100 simulations for each set of conditions.

**Cell transfection assays of miR-29 inhibitors and mimics**

Immature DCs from 4 unrelated individuals were transfected on day 5 using HiPerFect® transfection reagent. Cells were harvested and resuspended at $10^6$ cells/ml in complete medium without cytokines. 2 ml of cell suspension was added at once to 1ml of transfection medium (948 µl RPMI, 45 µl HiPerFect® and 4 µl of 20 µM oligonucleotide solution) in 6-well plates. After 6 h incubation at 37°C, 4 ml of complete medium containing 40ng/ml IL4 and 20ng/ml CSF2 was added to the cells and incubated at 37°C over night. miRCURY LNA Power Inhibitors were purchased from Exiqon (miR-29 family 460039, control 199020-00) and miRIDIAN microRNA mimics from Thermo Fisher (miR-29a C-300504-07, control CN-001000-01). At day 6, transfection efficiency was assessed by flow cytometry using a fluorescently labeled control oligonucleotide (Exiqon, 199020-04), and found to be on average 77% (Supplemental Fig. S11). Transfected cells were then infected for 24 h with MTB (H37Rv) as miR-29 induction peaks at this time (Supplemental Fig. S12).

**miR-29 quantification**

Total RNA was extracted using the miRNeasy kit (Qiagen). To quantify miR-29 expression upon MTB infection, cDNA was synthesized and quantitative real-time PCR (qPCR) performed using the Qiagen miScript PCR system and primers (miScript II RT Kit: 218161; miScript SYBR® Green PCR kit: 218073; miR-29a-3p MS00003262; miR-29b-3p MS00006566; miR-29c-3p MS00003269; U6 MS00033740). To validate miR-29 perturbation in transfected cells, cDNA was synthesized using the miRCURY LNA™ Universal cDNA Synthesis kit II (Exiqon, 203301) and qPCR performed using the ExiLENT SYBR® Green master mix and specific primers provided by Exiqon (miR-29a-3p 204698; miR-29b-3p 204679; miR-29c-3p 207429; RNU6-1 203907). qPCRs were

performed in a 7900 Real-time PCR system (Applied Biosystem). The relative expression of miR-29a, b and c, normalized to the endogenous control RNU6-1 was calculated using the $\Delta\Delta C_t$ method (Livak and Schmittgen 2001).

**Gene expression analysis**

Genome-wide profiling of non-infected and MTB-infected samples was obtained by hybridizing RNA to Illumina HumanHT-12 v4 Expression BeadChip arrays. RNA quality was assessed with the Agilent Bioanalyzer and all samples were of high integrity (RIN>8). Technical replicates were performed for 4 samples. Gene expression levels among technical replicates were highly correlated compared to biological replicates (mean Pearson's r = 0.997, $p< 1\times10^{-20}$), indicating high reproducibility. Initial microarray analysis was performed using the Bioconductor package lumi (Du et al. 2008). We performed a background correction and variance stabilizing transformation before quantile normalizing the data. We removed all probes that (i) did not map to a unique RefSeq ID, (ii) mapped to poorly characterized genes without an Ensembl ID, (iii) contained one or more HapMap SNPs (MAF>0.1 in the CEU population, i.e., Utah residents with ancestry from northern and western Europe) (The International HapMap Consortium 2010), or (iv) were detected in less than half of the samples. For genes represented by multiple probes, we used the mean probe expression level. These preprocessing steps yielded a final set of 12,722 probes, corresponding to 9610 genes, which were used for downstream analyses. Differential expression analysis was performed using the Bioconductor package limma (Smyth 2004) as described above. Enrichments of miR-29a predicted targets (Friedman et al. 2009) in differentially expressed genes were calculated using a Fisher's exact test. Differences in mean fold changes upon miR-29 perturbation between predicted targets, correlated genes identified by computational analyses and all genes were calculated using a t-test. Enrichments of functional Gene Ontology categories and KEGG pathways among differentially up- and down-regulated genes were computed using GeneTrail (Backes et al. 2007), as described above, using all detected genes as a background set.

**Quantification of cytokine and chemokine levels in supernatants**

Supernatants of untreated and MTB-infected DCs were 0.22μm filtered (Millipore) and kept at -80°C. We measured supernatant levels of 25 cytokines/chemokines, in duplicate, using the Human Cytokine Magnetic 25-Plex Panel (Invitrogen) according to manufacturer's instructions. We calculated the average quantity across technical replicates for each protein and used this for subsequent analyses. IL4 and CSF2 were excluded from the analysis as these cytokines were artificially added to the culture medium during DC derivation. CCL11 was removed from the

analysis as the median concentration in infected samples was lower than the detection limit specified by the manufacturer. Differences in secretion levels between conditions were calculated using a Wilcoxon paired rank sum test.

**References**


Backes C, Keller A, Kuentzer J, Kneissl B, Comtesse N, Elnakady YA, Muller R, Meese E, Lenhof HP. 2007. GeneTrail - advanced gene set enrichment analysis. *Nucleic Acids Res* **35**: W186-192.

Barreiro LB, Tailleux L, Pai AA, Gicquel B, Marioni JC, Gilad Y. 2012. Deciphering the genetic architecture of variation in the immune response to Mycobacterium tuberculosis infection. *Proc Natl Acad Sci U S A* **109**: 1204-1209.

Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J R Statist Soc B* **57**: 289-300.

Degner JF, Pai AA, Pique-Regi R, Veyrieras JB, Gaffney DJ, Pickrell JK, De Leon S, Michelini K, Lewellen N, Crawford GE, et al. 2012. DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature* **482**: 390-394.

Du P, Kibbe WA, Lin SM. 2008. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* **24**: 1547-1548.

Friedman RC, Farh KK, Burge CB, Bartel DP. 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* **19**: 92-105.

Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. 2006. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* **34**: D140-144.

Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* **106**: 9362-9367.

Howie BN, Donnelly P, Marchini J. 2009. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**: e1000529.

Irizarry RA, Wu Z, Jaffee HA. 2006. Comparison of Affymetrix GeneChip expression measures. *Bioinformatics* **22**: 789-794.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**: 402-408.

Lopez-Romero P. 2011. Pre-processing and differential expression analysis of Agilent microRNA arrays using the AgiMicroRna Bioconductor library. *BMC Genomics* **12**: 64.

Maranville JC, Luca F, Richards AL, Wen X, Witonsky DB, Baxter S, Stephens M, Di Rienzo A. 2011. Interactions between glucocorticoid treatment and cis-regulatory polymorphisms contribute to cellular response phenotypes. *PLoS Genet* **7**: e1002162.

Pickrell JK, Marioni JC, Pai AA, Degner JF, Engelhardt BE, Nkadori E, Veyrieras JB, Stephens M, Gilad Y, Pritchard JK. 2010. Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* **464**: 768-772.

Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ. 2010. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**: 2336-2337.

Smyth GK. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**: Article3.

The 1000 Genomes Project Consortium. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**: 56-65.

The ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57-74.

The International HapMap Consortium. 2010. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**: 52-58.

Xia K, Shabalin AA, Huang S, Madar V, Zhou YH, Wang W, Zou F, Sun W, Sullivan PF, Wright FA. 2012. seeQTL: a searchable database for human eQTLs. *Bioinformatics* **28**: 451-452.

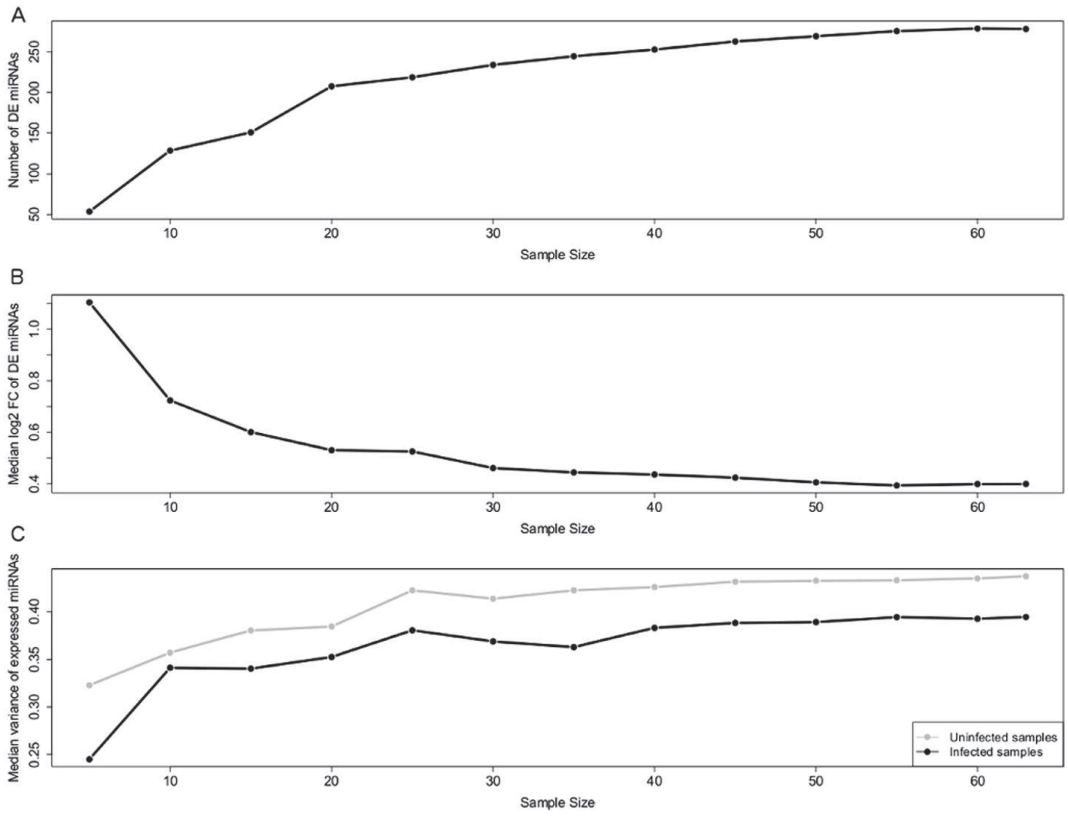



**Figure S1.** Relationship between sample size and differentially expressed miRNAs. Plots showing the changes in (*A*) mean number, (*B*) median $\log^2$ fold change and (*C*) median variance of differentially expressed miRNAs with increasing sample size. We assessed our resolution for differential expression analyses, compared to the smaller sample sizes more common for cellular studies, by repeating the analysis using randomly resampled subsets of 5 to 63 individuals, across 10 replicates. We observed an increase in the number of significantly differentially expressed miRNAs with larger sample sizes (*A*). At the same time, the average fold change of differentially expressed miRNAs decreased with increasing sample size (*B*), and the average variance increased (*C*). This demonstrates that larger sample sizes improve the detection of differential expression for miRNAs where the change is more subtle and/or inter-individual variation in expression more pronounced. Moreover, the plateau reached in all measures around 30 individuals indicates that the large number of differentially expressed miRNAs detected probably represents close to all changes occurring in miRNA expression profiles upon MTB infection.

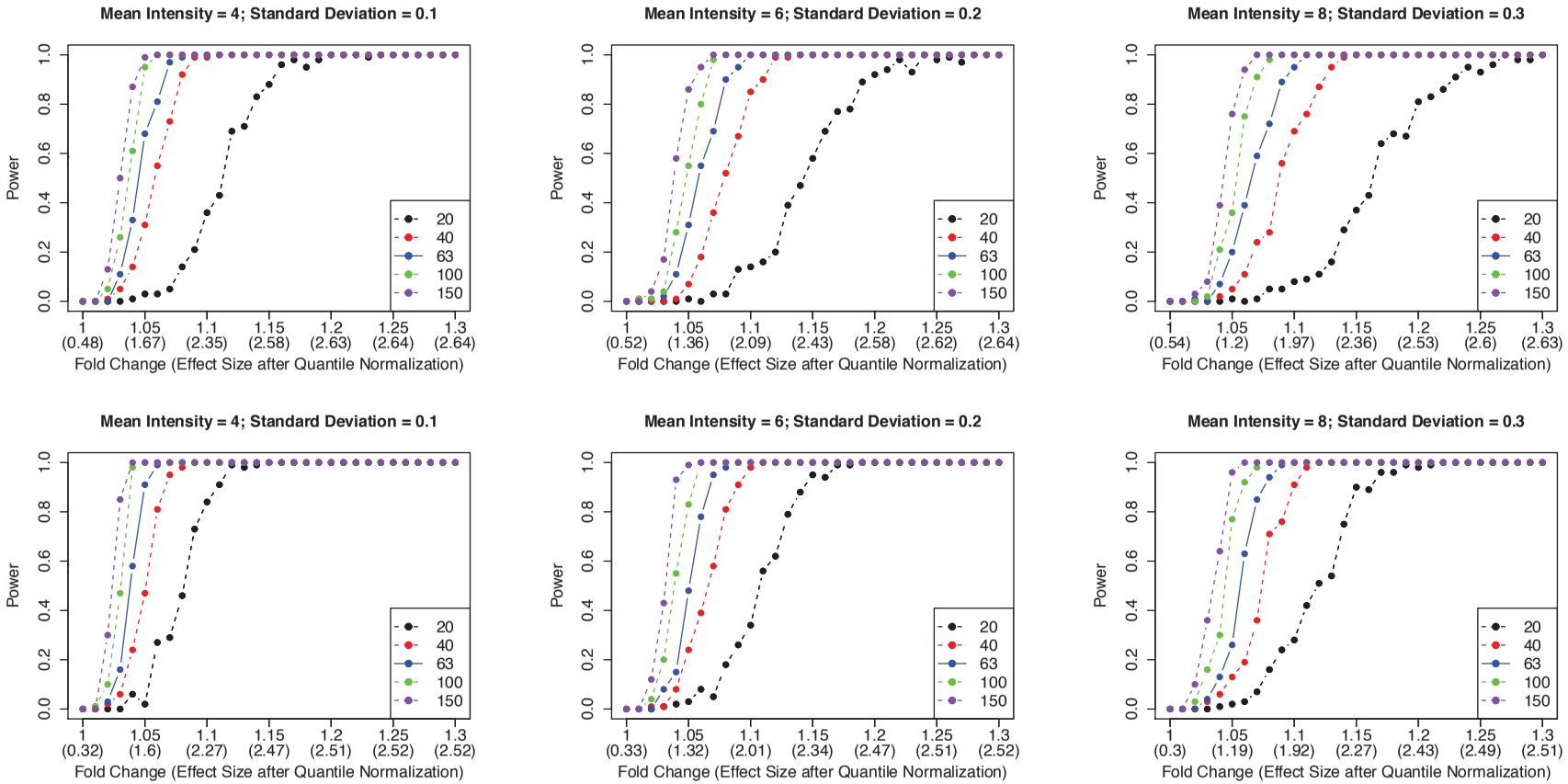



**Figure S2.** Power estimations for the detection of eQTLs. Plots showing the power to detect an association between a genotype and expression levels at $p<1\times10^{-4}$ for a range of sample sizes from 20-150 individuals, considering two different MAFs (0.2 (top) and 0.5 (bottom)) and three combinations of mean expression level and within genotype standard deviation in expression (4 and 0.1, 6 and 0.2, and 8 and 0.3). These values

correspond approximately to the 25$^{th}$, 50$^{th}$ and 75$^{th}$ percentiles of our observed values. We simulated fold changes between mean expression levels of the homozygote genotypes from 1 to 1.3. Effect size (given in brackets on the x-axis of the plots) was calculated as the standardized difference between the means of quantile normalized expression levels for the homozygote genotypes. The sample size of the current study (63 individuals) is shown with a solid blue line. We observed that the sample size used in this study gives reasonable power to detect associations when the effect size is greater than 1.5 and has almost 100% power when the effect size exceeds 2. In addition, the current sample size consistently performs almost as well as 100 individuals.

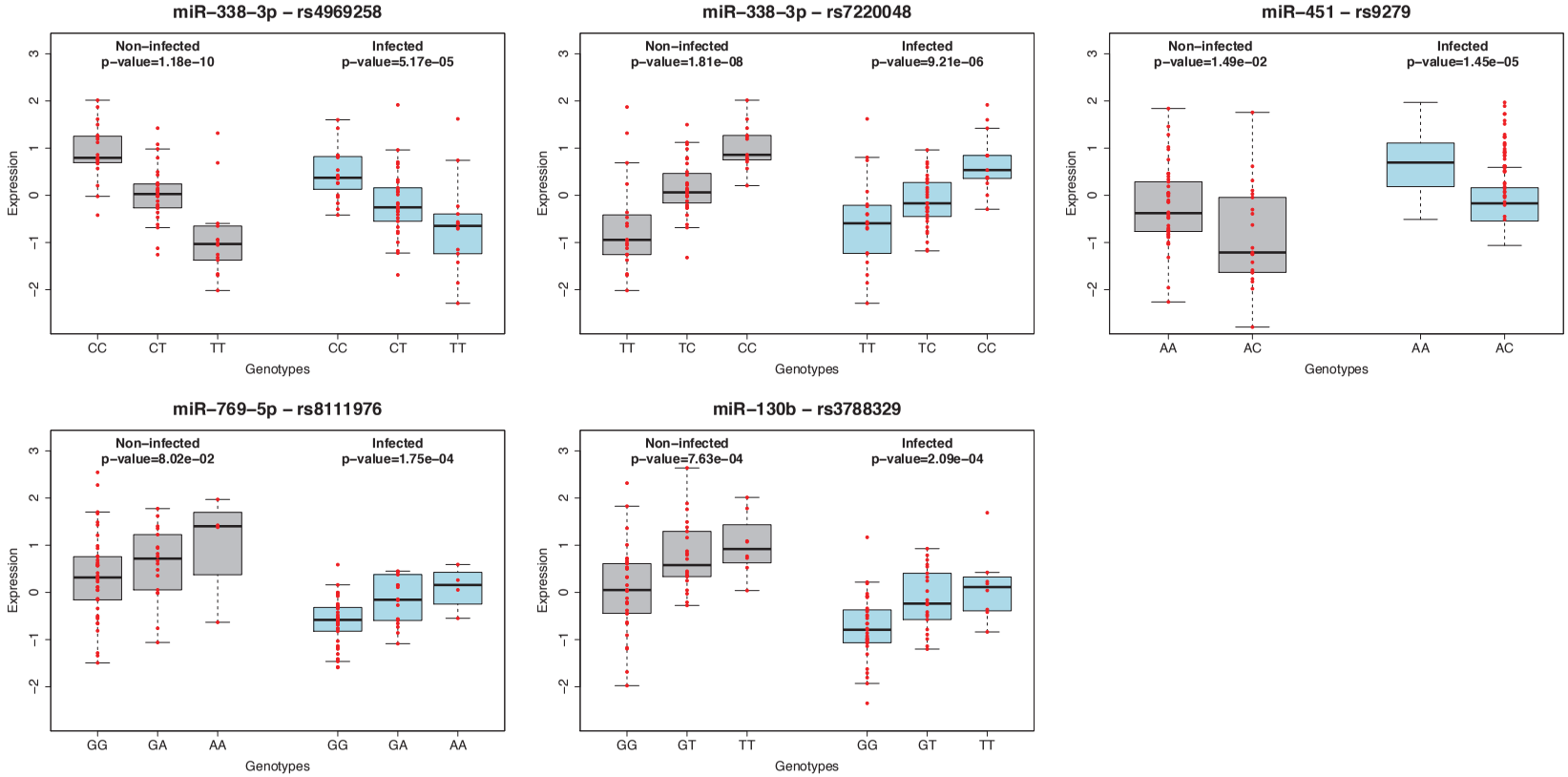



**Figure S3.** Boxplots of cis-eQTLs identified in this study. Boxplots showing all cis eQTLs identified in non-infected and/or infected DCs that did not satisfy the conditions to be considered as a response eQTL. Only one of these associations (miR-338-3p) reached genome-wide significance in both infected and non-infected samples, although different SNPs showed the strongest association with expression of the miRNA in each condition. BRIdGE analysis allowed us to refine this association by identifying an interaction effect for miR-338-3p/ rs4969258, where the association between the genotype and the molecular phenotype differs in its magnitude. All remaining associations were significant only upon MTB infection, although similar tendencies with regards to the effect of the genotype on miRNA expression before and after infection were observed.

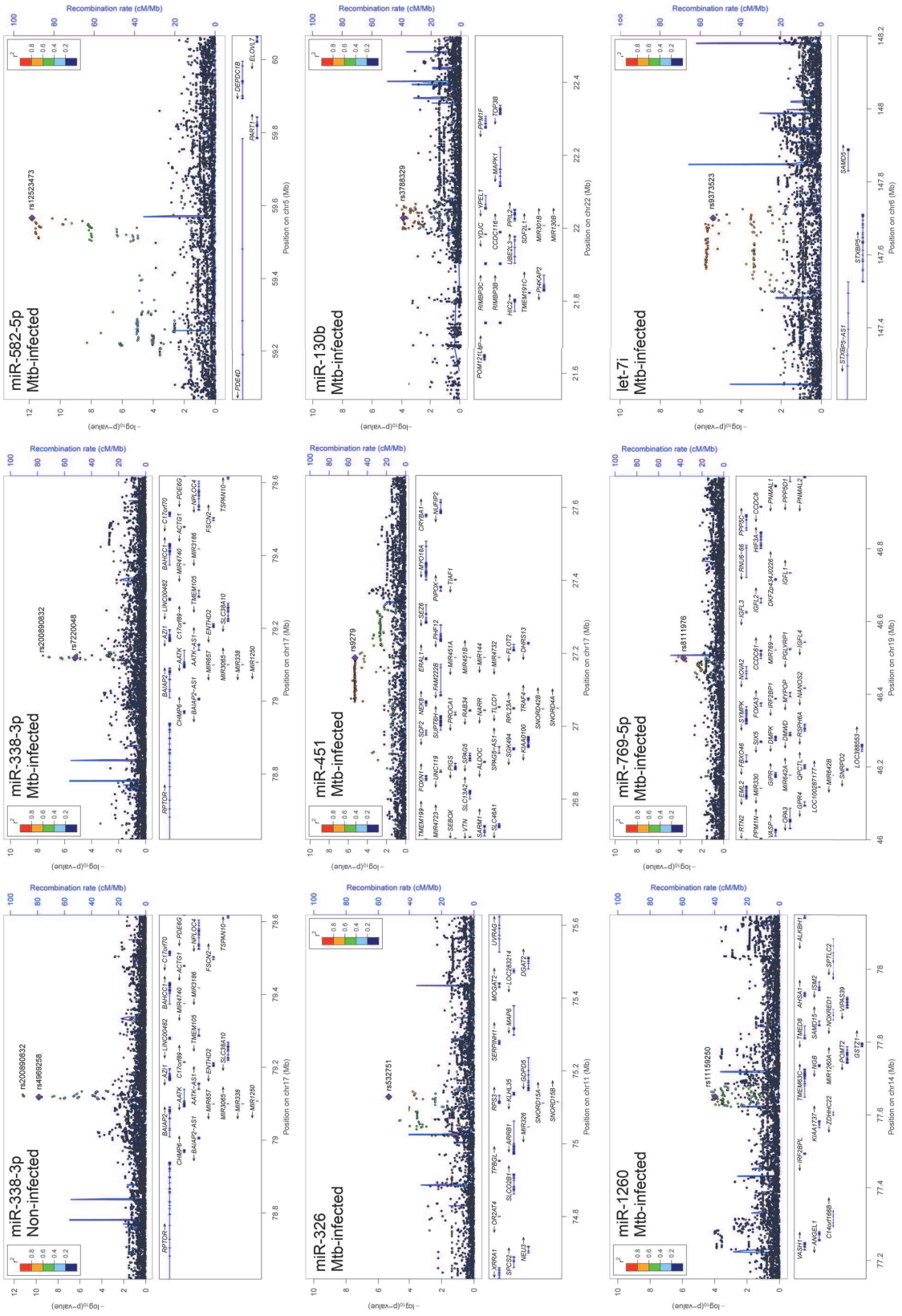

miR-582-5p
Mtb-infected
miR-130b
Mtb-infected
let-7i
Mtb-infected
miR-338-3p
Mtb-infected
miR-451
Mtb-infected
miR-769-5p
Mtb-infected
miR-338-3p
Non-infected
miR-326
Mtb-infected
miR-1260
Mtb-infected
Recombination rate (cM/Mb)
−log10(p−value)
Position on chr5 (Mb)
Position on chr22 (Mb)
Position on chr6 (Mb)
Position on chr17 (Mb)
Position on chr19 (Mb)
Position on chr11 (Mb)
Position on chr14 (Mb)
rs12523473
rs3788329
rs9373623
rs2008890832
rs7220048
rs9279
rs8111976
rs4969258
rs6532751
rs111592350

**Figure S4.** Regional association plots of genotyped and imputed SNPs surrounding detected miR-eQTLs. To refine our array-based miR-eQTL signals, SNPs that were not present on our array were imputed using data from the 1000 Genomes project, and eQTL mapping was repeated for all SNPs. Recombination rates and LD values are based on European-descent populations from the 1000 Genomes Project. The strongest array-based miR-eQTL SNP is named on the plot and denoted by a purple diamond. Grey circles indicate that no LD information is available for a given SNP. In almost all cases, the strongest array-based miR-eQTL SNP showed, or was in strong LD with the SNP showing, the strongest association with miRNA expression.

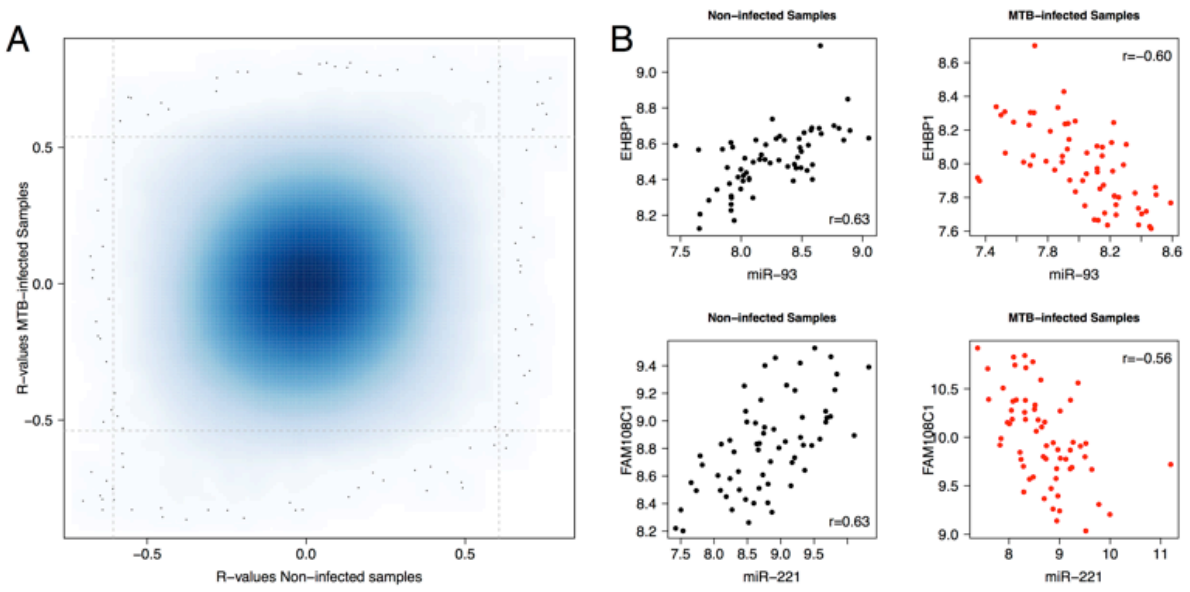



**Figure S5.** Impact of infection on miRNA-mRNA correlations. (*A*) Smooth scatter plot showing miRNA-mRNA correlations in non-infected and infected samples. Dashed lines represent the significance thresholds used to define the set of significantly correlated miRNA-mRNA pairs reported in the manuscript. Forty miRNA-mRNA pairs were significantly correlated in both analyses. The majority of this overlap was accounted for by correlations with 2 miRNAs – miR-155 and miR-210 – that together accounted for 88% of all overlapping pairs. Interestingly, while the vast majority of these 40 pairs showed the same direction of correlation before and after infection (r=0.88), 2 miRNA-mRNA pairs showed opposing correlations in the 2 conditions. (*B*) Scatter plots for these 2 miRNA-mRNA pairs, showing miRNA and mRNA expression levels for 63 individuals before and after infection. Both of these pairs were significantly positively correlated before and significantly negatively correlated after infection. As neither of these genes is predicted to be a direct target of the miRNA with which they are correlated, these changes are likely to reflect more general changes in the regulatory network upon infection.

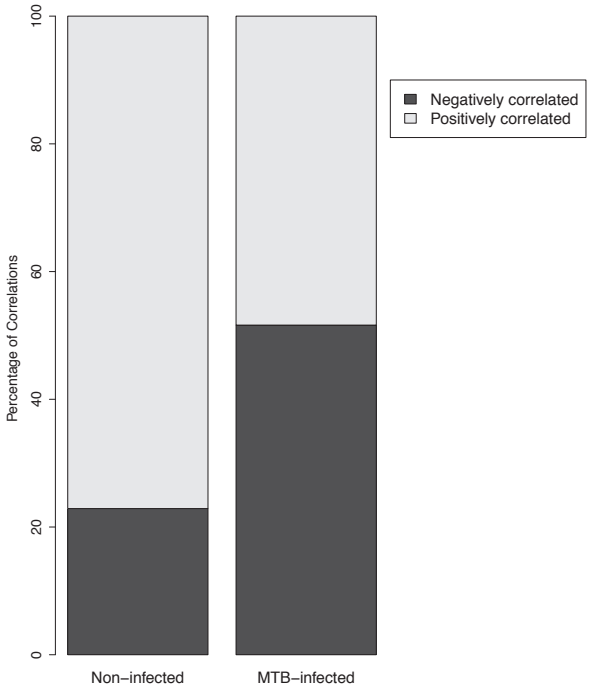



**Figure S6. Barplots of significantly correlated miRNA-mRNA pairs.** Barplots showing the proportions of negative and positive correlations among significantly correlated miRNA-mRNA pairs (FDR<0.005) in non-infected and infected samples.

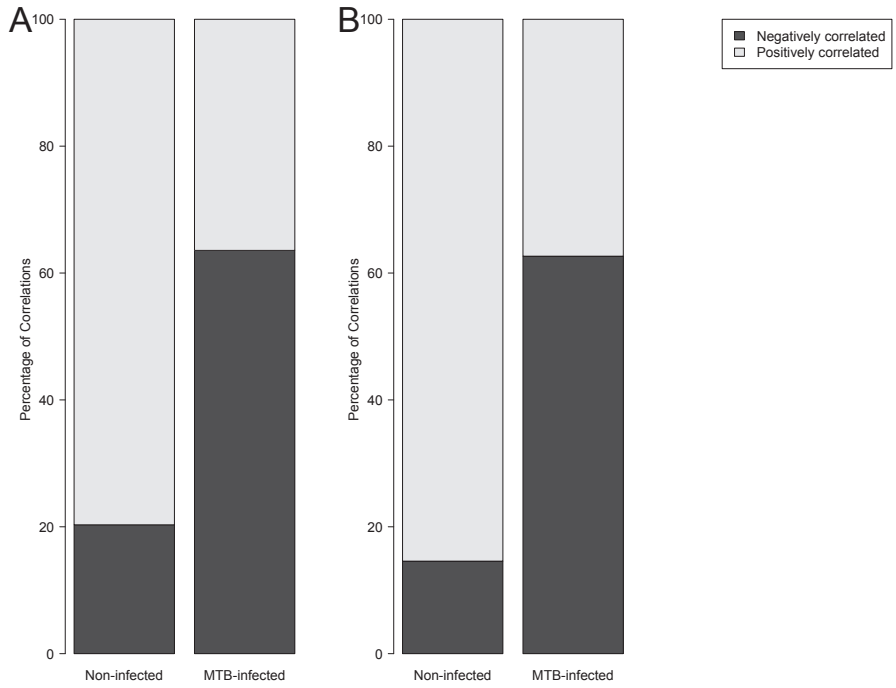



**Figure S7.** Variation in the percentage of infected cells and miRNA-mRNA correlations. Barplots of the genome-wide correlations between miRNAs and mRNAs (|r|>0.7) (*A*) before and (*B*) after correction for the percentage of infected cells. To assess the effect of inter-individual variability in the percentage of infected cells (ranging from 16-67%), we recalculated the miRNA-mRNA correlations for a subset of 47 samples for which information on the percentage of infected cells was quantified by FACS analysis using GFP-tagged bacteria. The effect of variation in the percentage of infected cells was corrected for using a linear model in non-infected (left panel) and infected (right panel) samples. In both analyses, at an |r|>0.7, we detected a greater number of significant correlations in non-infected samples, with respect to infected samples (not shown here). Moreover, among significantly correlated pairs, the majority of correlations in non-infected samples were positive, while after infection around 60% of miRNA-mRNA correlations were negative, consistent with the results obtained using the full data set. Furthermore, we observed a strong overlap in the significantly correlated miRNA-mRNA pairs across datasets (data not shown), suggesting that the percentage of infected cells does not have a strong influence on our results.

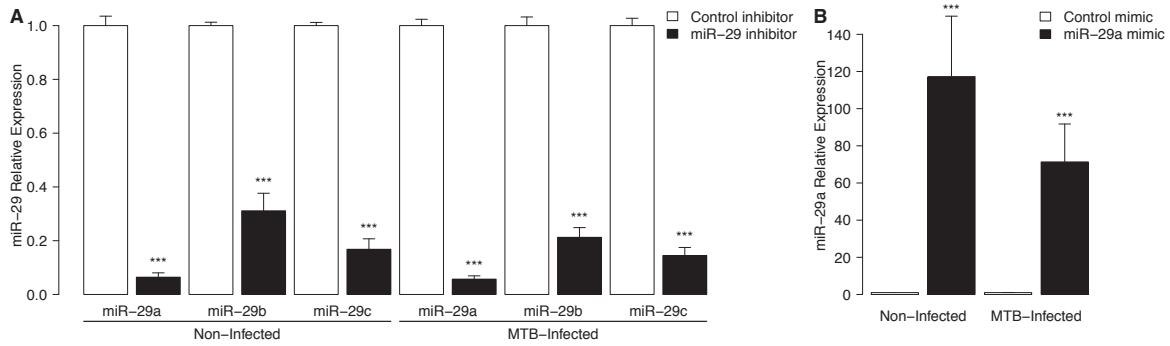



**Figure S8.** Perturbation of miR-29a expression using gain- and loss-of-function approaches. DCs were transfected with either (*A*) a miR-29 inhibitor or (*B*) a miR-29a mimic at day 5 and infected with MTB at day 6. Cells were lyzed and miR-29 levels quantified using qPCR, normalized on RNU6-1 levels. Fold expression was calculated with respect to miR-29 expression in control transfected DCs. The data represent the mean of duplicates of qPCR calculated across 4 different donors.

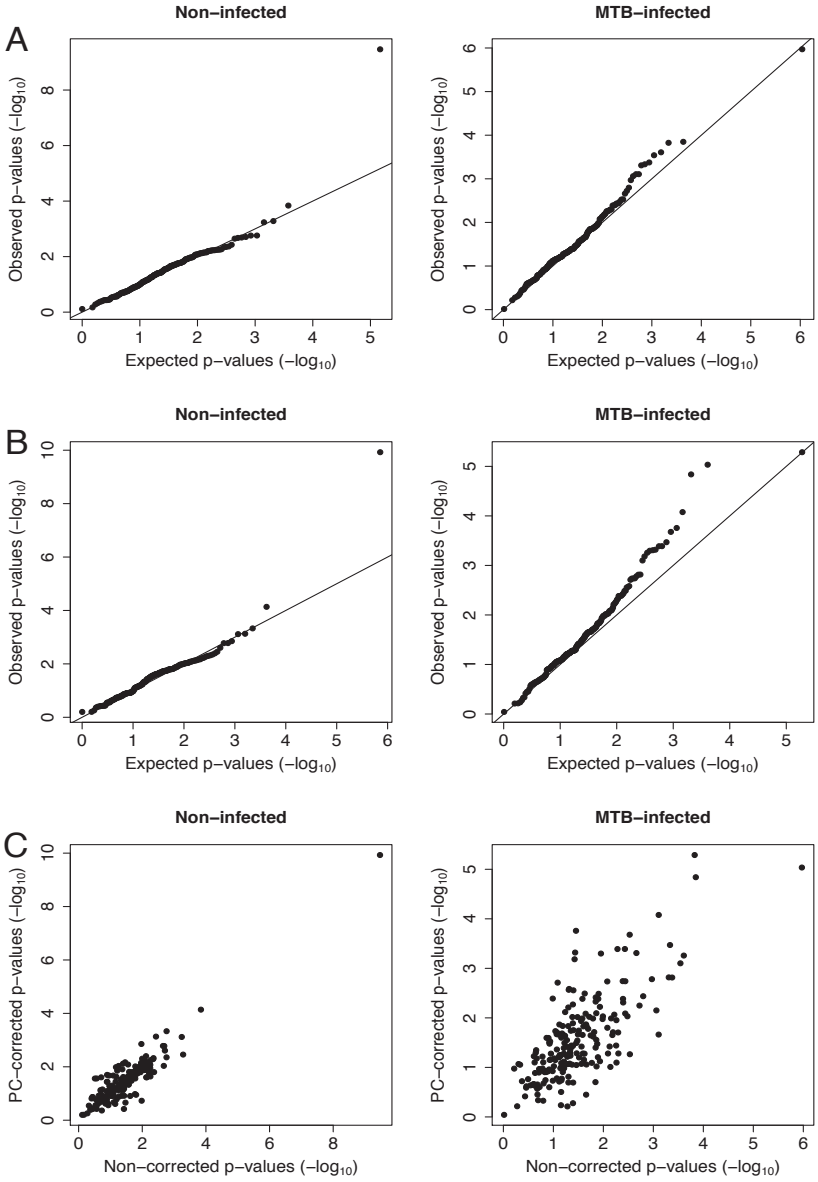



**Figure S9.** Principal component corrections for cis-eQTL detection. (*A*). QQ plots of minimum p-values for the association between miRNA expression and all SNPs within a 200kb window around the miRNA for non-infected (left panels) and infected (right panels) samples. Expected values were calculated based on 100 permutations of miRNA expression data. (*B*). QQ plots showing improved sensitivity of eQTL detection after accounting for unknown confounders using PC analysis. Specifically, the greatest number of significant associations was found after removing 1 and 4 PCs in non-infected and infected samples, respectively. (*C*). Plots showing the correlation between minimum p-values, before and after PC correction, indicating that this correction did not qualitatively change the relationships between genotypes and miRNA expression variation.

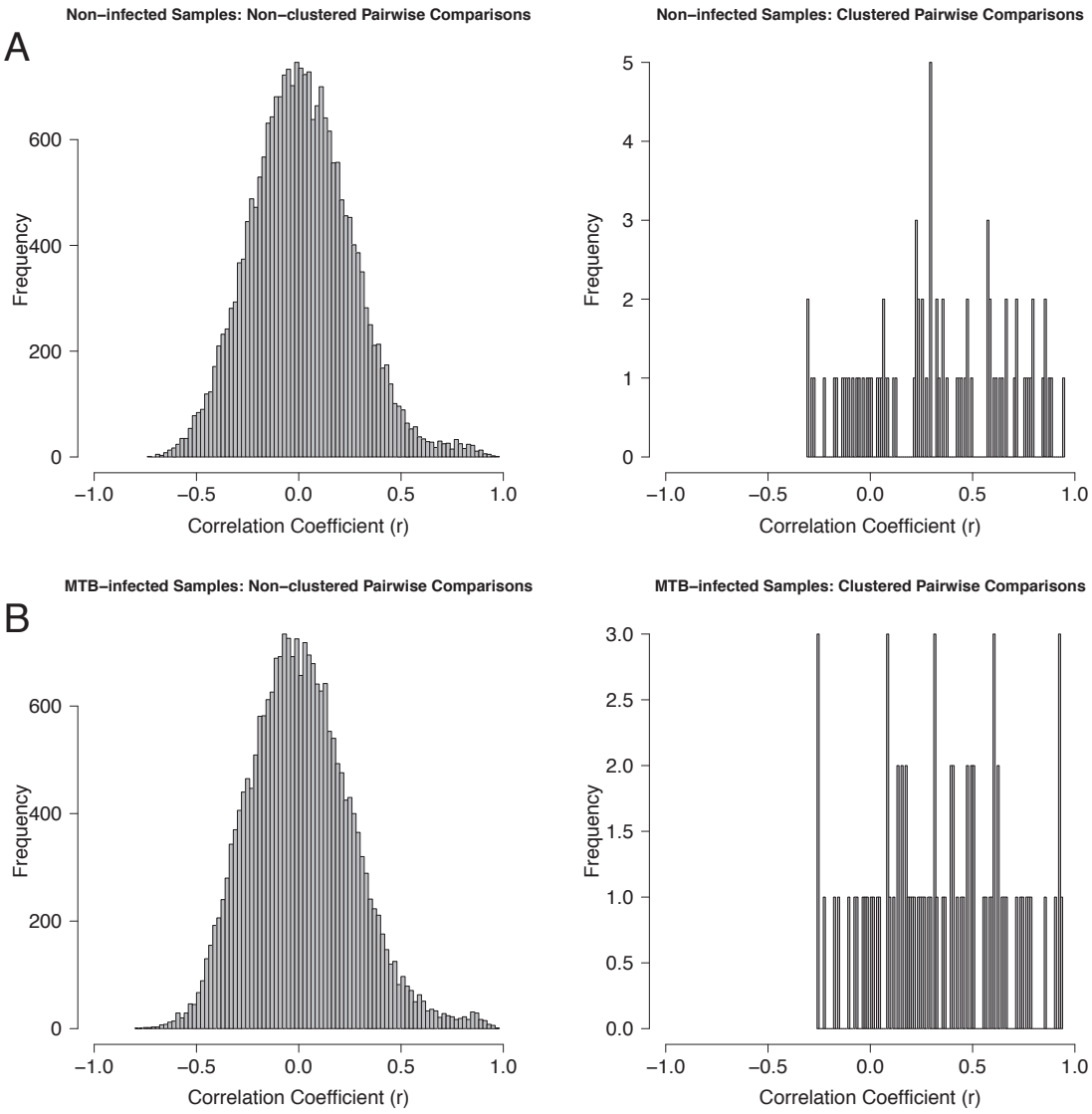



**Figure S10.** Distribution of correlation coefficients between miRNA pairs. We calculated the Pearson correlation coefficient pairwise among quantile normalized miRNA expression levels. Clustered miRNAs (right), those lying within 10kb of each other, were significantly enriched in positive correlations compared to the genome-wide distribution of correlations (left) in both (*A*) non-infected and (*B*) MTB-infected samples. This supports the hypothesis that independent mature miRNA sequences lying less than 10kb apart are frequently co-transcribed. This observation motivated our decision to consider only the most highly expressed transcript from each precursor in the analysis of miRNA-mRNA correlations, as this is the most likely to have an impact on mRNA expression levels.

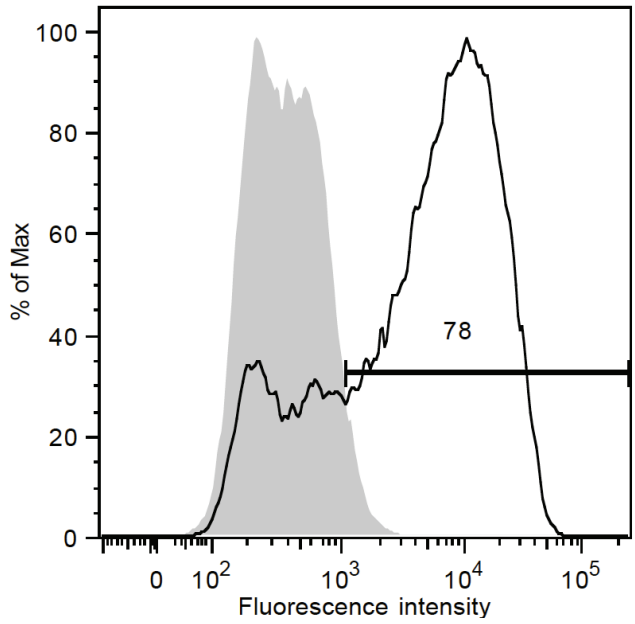



**Figure S11.** Transfection efficiency of DCs. DCs were transfected at day 5 using a fluorescently labeled negative control LNA Power inhibitor to evaluate the percentage of transfected cells (black). Fluorescence was compared to DCs transfected with an unlabeled control oligonucleotide (grey). Transfection efficiency was assessed by flow cytometry at day 6. Data shown are for one representative sample of four independent experiments performed on four different donors. The average transfection efficiency across all donors was 77%.

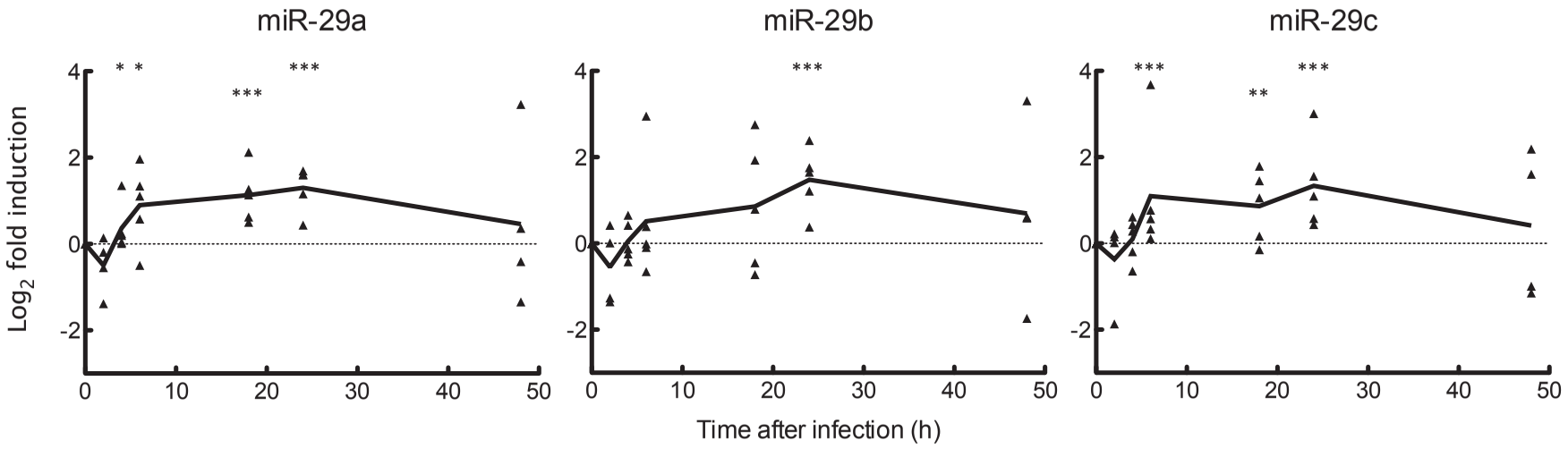



**Figure S12.** Time course of miR-29 family expression upon MTB infection in DCs. DCs were infected at day 6 with MTB (H37Rv) and lysed at six different time points (i.e., 2, 4, 6, 18, 24 and 48h) to quantify miR-29 expression levels. Expression levels of miR-29a, b and c were determined by qPCR, normalized on RNU6-1 levels, and fold induction was obtained by comparing to the non-infected condition. The data represent the mean of a duplicate qPCR calculated from at least 4 independent experiments, each derived from different donors. The significance of differences observed between MTB-infected and non-infected DCs was tested using a Mann-Whitney test (*$p<0.05$, **$p<0.01$, ***$p<0.001$).

## Résumé

Malgré les progrès médicaux réalisés au cours de ce dernier siècle, les maladies infectieuses demeurent l'une des principales causes de mortalité à travers le monde. La réponse immunitaire est l'un des phénotypes les plus complexes qui existent. Elle présente une certaine variabilité, tant au sein qu'entre les populations humaines. Cette thèse vise à identifier des facteurs génétiques et des mécanismes moléculaires sous-jacents aux différences de susceptibilité aux maladies infectieuses grâce à l'utilisation d'une combinaison d'approches *in silico* et *ex vivo*. Dans un premier temps, nous avons réalisé des analyses de génétique des population et de génétique évolutive en nous basant sur les données du projet 1 000 Génomes pour évaluer l'impact de la sélection naturelle sur les gènes de l'immunité innée. Nos résultats montrent l'étendue et l'hétérogénéité des pressions sélectives sur les gènes impliqués dans la réponse immédiate à l'infection. De plus, nos données suggèrent que l'introgression d'allèles provenant de l'Homme de Néandertal dans certaines de ces séquences comme le cluster *TLR6-1-10* et *SIRT1* ont participé à l'adaptation des populations Européennes et d'Asie Orientale aux pathogènes présents dans leurs environnements respectifs. Dans un second temps, nous avons estimé l'implication des miARN dans la réponse des cellules dendritiques à l'infection par *Mycobacterium tuberculosis*. Nos résultats soulignent les conséquences de l'infection sur les réseaux de régulation de l'expression des gènes par les miARN et montrent que l'expression de 3 % des miARN est associée à des facteurs génétiques localisés à proximité de ces gènes. Nous identifions en particulier deux associations concernant miR-326 et miR-1260 qui ne sont observées que dans un contexte infectieux. Le travail présenté ici constitue la plus large étude de génétique évolutive et de génétique des populations axée sur les gènes de l'immunité innée réalisée à ce jour et la première caractérisation de l'architecture génétique de la réponse à l'infection impliquant les miARN. Notre travail participe à l'amélioration des connaissances que nous avons sur les bases génétiques de la variabilité de réponse immunitaire chez l'Homme et propose un ensemble de variations génétiques à caractériser fonctionnellement dans le but de révéler de nouveaux mécanismes moléculaires sous-jacents aux différences de susceptibilité aux maladies infectieuses.



## Abstract

Despite the major medical advances of the last century, infectious diseases remain one of the leading causes of death worldwide. The immune response to pathogens is one of the most complex phenotypes that exist and presents substantial variability among individuals and populations. This thesis aims to identify genetic factors and molecular mechanisms underlying differences in susceptibility to infectious diseases using a combination of *in silico* and *ex vivo* approaches. First, we performed population and evolutionary genetics analyses using the 1,000 Genomes Project dataset to assess the impact of natural selection on innate immunity genes. Our analyses reveal the widespread and heterogeneous nature of the selective pressures acting on genes involved in innate immune processes in humans. In addition, we suggest that the introgression of Neanderthal alleles in some innate immunity genes, especially the *TLR6-1-10* cluster and *SIRT1*, contributed to the adaptation of European and East Asian populations to local pathogens. Second, we profiled the miRNA response to *Mycobacterium tuberculosis* infection in human dendritic cells. Our results highlight the impact of infection on miRNA-mediated gene regulatory networks and show that the expression of 3% of miRNAs is associated with proximate genetic variants. More specifically, we identify two infection-specific associations for miR-326 and miR-1260. The work presented here provides the largest evolutionary genetics analysis of innate immunity genes to date and the first attempt to characterize the genetic architecture of the miRNA response to infection. Our work offers new insights into the genetic basis of inter-individual variability in immune responses and provides a set of candidate genetic variants for future functional validation to elucidate novel molecular mechanisms underlying differences in susceptibility to infectious diseases.